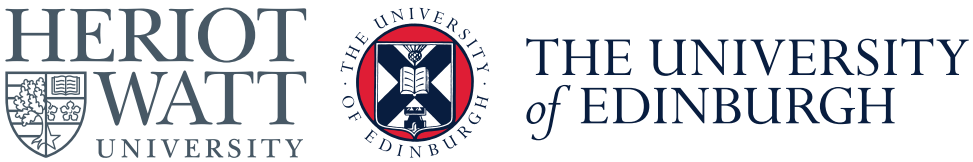
State Space Reparametrization For Approximating Nonlinear Models In Bayesian State Estimation

José Luis Franco Monsalve

A Thesis Submitted for the Degree of
Doctor of Philosophy in Electrical Engineering



February 2017

Abstract

Recursive Bayesian state estimation is a powerful methodology which is useful for the integration of data about a process of interest while considering all the sources of uncertainty which are present in the observations and in modeling inaccuracies. However, in its general form it is intractable and approximations need to be made in order to use it in real life applications. The most widely used algorithm to perform recursive state estimation is the Kalman filter, which assumes that the probability distributions that it propagates are Gaussian and that the measurement and dynamical processes are linear. If these assumptions are satisfied, the Kalman filter is optimal. In most applications, however, this proves to be an oversimplification, due to which several techniques have arisen to handle model non-linearity and different types of distributions.

In this thesis, a novel method for the estimation of distributions with nonlinear dynamical and measurement models is presented, which uses a reparametrization of the state space of the distributions in order to exploit the linear properties of the Kalman filter. This involves the mapping of the distribution into a different space, and a subsequent approximation as a Gaussian distribution. An analysis of the adequacy of this transformation is presented, which shows that it is a valid approach in a number of practically interesting filtering problems.

The proposed approach is applied to the estimation of the state of Earth-orbiting objects, as it is a challenging estimation scenario which can benefit from the use of filter. Space situational awareness is increasingly important as near-Earth space becomes cluttered with satellites and debris. In this work, the sensors that are most commonly used to track objects in orbit, radars and telescopes, are modeled and a filter based on the previously discussed ideas is proposed.

Finally, a multi-object estimation filter based on a recent estimation framework is presented which propagates high amounts of information while maintaining low computational complexity. This is important as there are many challenges to tracking large amounts of orbiting objects in a principled way using ground-based sensors, and naturally extends the single object filter described above to the multi-sensor, multi-object case.

Contents

List of Figures

List of Algorithms

List of Tables

Acknowledgments

Producing this PhD. thesis has been the fruit of a large effort, which I would not have been able to manage without the help of many others. I am grateful to my family for their support in setting me on this journey, and to the people who have inspired me to be curious, whether by teaching or by example. I am indebted to the European Commission for the Vibot Erasmus Mundus scholarship without which I would surely not have started this PhD., and to the University and UDRC for providing me with the financial support which enabled me to carry it out.

I have had the fortune to work with very talented people. Thanks to Daniel for being a very engaged supervisor, and leading a team with such high standards of research; It was due to his trust in me that I decided to start this program. Thanks to Manu and Jérémie for setting an example of hard work, rigor and brilliance which have been an inspiration throughout, and to my fellow students and collaborators for their engagement in producing very high quality work.

I am tremendously fortunate to have met an incredible group of friends in Edinburgh, and I am very thankful to them for having made my stay in this city a pleasure. I cannot mention everyone here, but I want to give special thanks to some of them. Thanks to Liam for being contagiously curious and positive, to Shailendra for the good times in so many pubs in Edinburgh, to Peter for the music and for the clever conversations, to Nico for reminding me to take a step back every now and then and breathe. Thanks to Gena, Sarthak and Manu for being not only great friends but also very talented brewing partners, to Corina for being such a good flatmate, to José for all the good times, to Pablo for showing me so many opportunities and pushing me to go after them. Thanks to Alexey for the great times, so many discoveries, and for the time he spent proofreading this work. And thanks to Selma for so much warmth, for all the support, and for being such an inspiration.

— A riddle, sir? Ask me, sir.
— O, ask me, sir.
— A hard one, sir.
— This is the riddle, Stephen said:

The cock crew,
The sky was blue:
The bells in heaven
Were striking eleven.
'Tis time for this poor soul
To go to heaven.

What is that?

— What, sir?
— Again, sir. We didn't hear.
Their eyes grew bigger as the lines were repeated. After a silence Cochrane said:
— What is it, sir? We give it up.
Stephen, his throat itching, answered:
— The fox burying his grandmother under a hollybush.

Ulysses
James Joyce

Chapter 1

Introduction

STATISTICAL state estimation is an area of great interest to the engineering community. With the advent of machines that needed to be controlled based on their state, a reliable means of estimating said state was required in order to take appropriate control actions. The scope of state estimation, however, goes far beyond control applications. Indeed, it is an essential tool in disciplines such as defense, finance, robotics and many more, as it provides a reliable way of handling uncertainty in problems where measurements can not be assumed to give perfect information on the observed system.

The need to estimate the state of increasingly complex systems, coupled with the rapidly expanding availability of low-cost, powerful computer hardware, has spurred the development of estimation algorithms of growing sophistication. The advent of the Kalman filter in the sixties was essential for the success of the Apollo missions, as it permitted to offset the navigation errors with a continual stream of measurements from their on-board sensors [30]. More recently, sophisticated particle filtering algorithms have been fundamental for the development of autonomous vehicles whether in air, on the ground, or underwater [94].

Using probability distributions rather than point estimates gives a considerably large amount of information which can be used when decisions need to be taken based on knowledge about the estimated process. Faithfully representing uncertainty in a diversity of domains, however, is a challenging task due to the variety of dynamical and measurement models that arise from different application areas. Filtering techniques such as the Kalman filter [47] have proved to be useful in situations where distributions can be reasonably represented with a Gaussian approximation, and the modeled systems are simple enough to be represented with linear transformations. They also benefit from not requiring a large amount of computational resources,

since they use lean representations for the probability distributions. In more complex systems, however, these approximations can result in undesirable effects such as biased distributions, over or underconfident estimations of the uncertainty, and more. More complex filters will naturally require more computational resources to overcome these problems.

A central subject in this thesis is the combination of linear and nonlinear techniques in order to minimize the additional resource requirements while still faithfully propagating the distributions. This will usually be carried out by reparametrizing the distributions in order to apply linear operations to the distributions, circumventing the problems associated with nonlinear dynamical or observation models.

Although estimating the state of single targets has been widely studied, a more general problem is estimating the state of *populations* of objects, where both the number of targets and their states are unknown [5]. This problem is not as straightforward as running a number of single target filters in parallel, as it is also necessary to keep in mind the presence of false alarms, where spurious measurements not due to any object of interest in the population appear in the collected data; missed detections, as the probability of detecting individuals in the scene may be less than unity; and data association, as it is necessary to evaluate which measurement correspond to which object in the scene.

Classical approaches to multiple object estimation are based on heuristic extensions to single target filters, where data association techniques based on statistical distance are used to hypothesize possible assignments from measurements to tracks, misdetections, or false alarms. Although these approaches are widely used, the heuristics that are used for track management introduce the need for parameter tuning and make it impossible to verify the validity of the filters, e.g. by providing convergence results. A more recent formulation for multiple object state estimation, called finite set statistics [58], is based on the concept of propagating probability distributions based on random sets rather than on random vectors, which has produced principled filters without the previously mentioned limitations of the classical approach. Due to the set representation, however, track identity is not directly propagated between time steps.

A new approach for estimating *stochastic populations* [37] has been recently conceived which attempts to combine the best aspects of both classical approaches, where track identity is preserved, and finite set statistics, which provides a principled, extensible framework. Based on the concept of distinguishability, this very recent approach has already produced filters which outperform their state-of-the-art equivalents, especially in scenarios with particularly low probabilities of detection,

while remaining fully probabilistic in the sense that they do not require the use of heuristics at any stage of the filtering process.

Using principled multi-object state estimation filters has the advantage that since they represent multi-object probability distributions, it is simple to manipulate them in order to create more complex estimation systems. For instance, sensor calibration or localization can be achieved by modeling the joint sensor-population distribution as a hierarchical point process. Multi-sensor fusion can be done through successive updates on the probability distribution using different sensors, and target classification can be done by using multi-model state representations.

This work is principally motivated by the increasing need to have a clear picture of near-Earth space, as the relevance of space-based infrastructure is essential for areas like communications, localization, and defense, among others, and their safety is endangered by the growing amounts of debris residing in orbit. Orbiting objects are typically observed using ground-based sensors such as radars and telescopes, and catalogs of their state is maintained by several agencies. The main available catalogs, however, do not include measures of uncertainty in their estimates, which is essential to evaluating important quantities such as the risk of collision for important assets, or the imprecision in measurements obtained from satellite-mounted sensors that is induced by the uncertainty in their position.

The final goal of this thesis is to present a method which provides a fully probabilistic view of near-Earth space, based on noisy measurements from ground based sensors, and which is able to adapt to observing previously undetected objects or sensor clutter. In order to do this, several aspects of estimation for orbital objects will need to be studied. A way to estimate the probability distribution of newly measured objects based on observations that do not include their full state will need to be proposed – telescopes, for instance, cannot measure the distance to an object from Earth and thus cannot resolve its full position. Also, it will be necessary to analyze the dynamics of orbiting objects and how they can be cast in a probabilistic framework, alongside modeling ground-based sensors and the uncertainty that they introduce. Finally, a suitable multi-object estimation framework will need to be implemented in order to manage different tracks and give individual information about observed objects in the presence of data association ambiguity, missed detections and false alarms.

In this context, several important challenges arise. The dynamics of orbiting objects are known with a high degree of accuracy, meaning that there is very low uncertainty in the dynamical models, complicating the use of Monte Carlo methods to propagate distributions. At the same time, their non-linearity makes it challeng-

ing to use Kalman filter-like algorithms for estimation. In this work, a method which attempts to use the best of both filtering paradigms will be proposed to tackle this problem. In terms of the estimation of populations, it is desirable to have an efficient method which scales well with the number of targets, as the amount of objects to track in orbit is very large.

In this thesis, the previously mentioned filtering principles are illustrated in this domain, and it is shown how they can be used to robustly and efficiently estimate the state of orbiting objects alongside their uncertainty, and maintain a catalog which takes into account the shortcomings of the sensors and the incomplete knowledge of object dynamics.

This manuscript is divided as follows. In Chapter 2, the background of Bayesian estimation and space situational awareness are discussed. Chapter 3 describes a single-object filter for estimating the state of objects in orbit, and Chapter 4 extends this to the multi-object case using a novel filtering method. Chapter 5 concludes the thesis. Appendix A describes statistical tests for multivariate normality, and Appendix **??** provides an application to estimation from video data which motivates the reparametrization approach presented in this thesis.

1.1 Objectives

The objectives of this thesis are the following:

- Present the Bayesian framework for state estimation, alongside commonly used filters that are derived from it, for single and multiple objects.
- Introduce a method to exploit linear filtering techniques in non-linear problems by reparametrizing the state distribution into the sensor space.
- Show a new algorithm for estimating the state of objects in orbit, from data obtained through radio-frequency and electro-optical sensors.
- Extend the previous algorithm to track multiple objects in orbit using a novel multiple object estimation framework.

1.2 Contributions

During the course of this PhD project, the following publications have been produced with contributions from the author, with a note about the level of involvement in the particular work:

- J. Franco, J. Houssineau, D. E. Clark, and C. Rickman. Simultaneous tracking of multiple particles and sensor position estimation in fluorescence microscopy images. In *Control, Automation and Information Sciences (ICCAIS), 2013 International Conference on*, pages 122–127. IEEE, 2013 — Principal author.
- J. Franco, E. D. Delande, C. Frueh, J. Houssineau, and D. E. Clark. A spherical co-ordinate space parameterisation for orbit estimation. In *Proceedings of the 2016 IEEE Aerospace Conference*, pages 1–12, 2016 — Principal author, this is the basis for Chapter 3.
- J. Franco, E. D. Delande, C. Frueh, J. Joussineau, and D. Clark. Probabilistic orbit determination using a sensor co-ordinate parametrization. *Journal of Guidance, Control and Dynamics*, —(—), Under review — Principal author, expands upon the previous publication.
- C. S. Lee, J. Franco, J. Houssineau, and D. E. Clark. Accelerating the single cluster PHD filter with a GPU implementation. In *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on*, pages 53–58. IEEE, Dec 2014 — Experiments.

- I. Schlangen, J. Franco, J. Houssineau, W. T. E. Pitkeathly, D. E. Clark, I. Smal, and C. Rickman. Marker-less stage drift correction in super-resolution microscopy using the single-cluster PHD filter. *IEEE Journal of Selected Topics in Signal Processing*, 10(1):193–202, 2016 — Concept, writing, part of the experiments.

- J. Houssineau, D. E. Clark, S. Ivekovic, C. S. Lee, and J. Franco. A unified approach for multi-object triangulation, tracking and camera calibration. *IEEE Transactions on Signal Processing*, 64(11):2934–2948, 2016 — Experiments, validation of the distributions. This publication inspired the development of the methods shown in Chapter 3.

- O. Hagen, J. Houssineau, I. Schlangen, E. D. Delande, J. Franco, and D. E. Clark. Joint estimation of telescope drift and space object tracking. In *Aerospace Conference, 2016 IEEE*, pages 1–10. IEEE, 2016 — Support with the basic algorithm that was used in the application.

- C. Simpson, A. Hunter, S. Vorgul, E. D. Delande, J. Franco, and D. E. Clark. Likelihood modelling of the space geodesy facility laser ranging sensor for Bayesian filtering. In *Sensor Signal Processing for Defence (SSPD), 2016*, pages 1–5. IEEE, 2016 — Concept, writing, supervision of the students. Case study in Chapter 2.

- A. Pak, J. Correa, M. Adams, D. E. Clark, E. D. Delande, J. Houssineau, J. Franco, and C. Frueh. Joint target detection and tracking filter for Chilbolton advanced meteorological radar data processing. In *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2016 — Support with basic concepts.

- E. D. Delande, J. Houssineau, J. Franco, C. Frueh, and D. E. Clark. A new multi-target tracking algorithm for a large number of orbiting objects. In *Proceedings of the 27th AAS/AIAA Space Flight Mechanics Meeting*, San Antonio, TX, 2017 — Implementation, integration of SSA models, experiments, writing. Forms the basis for Chapter 4.

- E. D. Delande, C. Frueh, J. Franco, J. Houssineau, and D. E. Clark. A novel multi-object filtering approach for space situational awareness. *Journal of Guidance, Control, and Dynamics*, 2017. submitted — Integration of SSA models, experiments.

The research project which the author was part of involved work from many collaborators. The particular contributions of the author were the following:

Chapter 2 Performed literature review. Supervised students in the development of range-tracking filter for Herstmonceux laser ranging facility.

Chapter 3 Developed probabilistic initial orbit determination method for radar and optical sensors. Collaborated in the creation of a radar tracking filter and developed the optical filter. Validated the distributions using Henze-Zirkler tests. Performed experiments and wrote two articles based on the work in this chapter.

Chapter 4 Implementation and integration of SSA models into the HISP filter. Experiments and validation.

Appendix A Literature review, implementation of test for different scenarios.

Chapter 2

Background

This chapter covers probabilistic state estimation from a Bayesian point of view; that is, the integration of observation data from a process of interest with models about how this data is generated and how the process behaves in order to produce probability distributions that describe its state. This is framed in the context of Space Situational Awareness (SSA), as it will be the main application focus for this thesis. First, an overview of techniques used for estimation in space situational awareness will be presented, followed by a detailed description of the general Bayesian filtering paradigm. After this, commonly used filters for the estimation of the state of single objects will be presented, after which a review of state estimation methods for populations of objects where not only the number of objects in the population must be estimated, but also their individual states, will be made. Finally, methods which extend these estimation methods in order to additionally estimate parameters of the observed process will be described. This will give an overview of the state of the art of state estimation, which will set the stage for the contributions of the thesis in later chapters.

2.1 Estimation in space situational awareness

Space infrastructure plays an increasingly important part in modern communications, reconnaissance, and geolocation, among other applications, and as more nations increase their stake in the exploitation of near-Earth space the number of artificial satellite launches increases year after year. Each of these launches generates debris which endangers current and future missions, and in spite of mitigation efforts this remains a very relevant problem. Safeguarding orbital assets, then, involves knowing their position and velocity with a high degree of accuracy and pre-

cision, and also any potential collision risks that could be posed by other satellites or by space debris [49].

Using noisy data to estimate the state of objects in orbit can be accomplished in several ways. For instance, when a fixed amount of data is available to estimate the parameters of an object, nonlinear curve fitting algorithms can be used [32]. Applications for this include the estimation of orbital parameters in remote planetary sensing applications [92]. In this thesis, the recursive Bayesian estimation paradigm is employed to estimate the state of orbiting objects as it enables the integration of data as it comes, giving instantaneous information about the uncertainty of the orbital estimate. This is particularly valuable when unknown objects appear in orbit, or there are sudden perturbations which change the state of known objects. Additionally, it enables the implementation of tractable multi-object estimation algorithms as data association only needs to be performed from a time step to the next rather than over all available data in past time instants.

Since the very beginning, the Kalman filter and its extensions have been invaluable in the domain of space situational awareness. From its use to plan and execute the Apollo moon missions [30] to global navigation systems such as GPS [31], passing by orbit determination and re-entry estimation methods [43], it has been indispensable in most developments which have enabled humankind to explore space.

Although techniques to propagate orbits from a known initial state have been widely studied, the problem of estimating the collision risk of two objects naturally benefits from knowing what the uncertainty of its position and velocity are. A problem with using only deterministic propagation to predict the position of an object is that the object may drift away from its initially estimated orbit due to perturbations such as space weather effects. Additionally, if an object is only observed once, unique orbit determination is not possible as only a subset of the full state is observable [95]. With this in mind, filtering algorithms such as the Kalman filter and its extensions attempt to propagate *probability distributions* rather than only a point estimate, and to use incoming measurements to decrease the uncertainty of the estimated orbit [8].

The most commonly used sensors used for space situational awareness are radar and optical sensors [76], although other sensors can be used such as laser ranging systems [86]. Radars are commonly used to track objects in space, and combined measurements typically give information about the azimuth and elevation of the object, its distance to the station, and the rate of change of this distance when Doppler information is available. Combined optical measurements are obtained from

telescopes, and give azimuth, elevation, and their rates of change. While telescopes can see objects that are very far away, they rely on passive illumination by the sun and clear weather conditions; whereas although radars have the advantage of using active sensing, it is costly to see objects that are far since radar energy returns are inversely proportional to the fourth power of the distance [76].

When the initial state of an object that is being observed is not known, it can be recovered deterministically if three measurements from the same target are available using Gauss' method, double R iteration [21] or Gooding's method [27]. This is prone to errors due to measurement noise, however, and requires a reliable method to determine that the measurements come from the same target. A recent development in orbit determination has produced the admissible regions approach, which offers constraints on the possible states of a target given that a single measurement is available [95]. In the Bayesian context, it is possible to use these energy constraints to generate a prior distribution based on a single measurement. This will be elaborated in chapter 3.

A limitation of the Kalman filter and its nonlinear extensions is that it relies on linearization of the orbital dynamics and observation models to produce its estimates, which will fail if the estimate uncertainty is too large causing the linearization to lose validity [45]. This is particularly the case for the distributions of objects that have been observed only once, as the range of values, and thus the uncertainty, of the unobserved parameters is very large [95].

Classical approaches to solve the statistical orbit determination problem rely on the Extended Kalman Filter [8] and its variants, which can only represent Gaussian distributions, and thus the banana-shaped uncertainties that are found in orbital estimation problems cannot be represented by it. This problem has been recognized by the community, and solutions based on Gaussian sum filters have been proposed instead, which are more flexible in the representation of the distribution [93, 96]. Another issue is that the measurement models are highly nonlinear so important information can be lost through their linearization; and as it was shown in [45], it gives biased estimates for range-bearing style problems as is the case in orbit determination. For this reason, the Unscented Kalman Filter [45] has been explored in orbital estimation situations. It relies on the propagation of the first two moments of the filtering distribution, which are however insufficient to represent arbitrary priors.

The bootstrap filter [29] is commonly used in problems where the dynamical or measurement models are nonlinear, and allows for modeling of arbitrarily shaped distributions. The performance of particle filtering with respect to classical approaches

has been demonstrated in [60], and hybrid approaches which also use a UKF when measurements are available have also been explored to reduce the uncertainty when measurements are acquired [70]. Another way of representing uncertainty, based on generalized polynomial chaos, is able to model parametric uncertainty in addition to perturbations and uncertain initial conditions [51].

Several approaches have been used in the past to track multiple objects in space. The MHT [87], Labeled Multi-Bernoulli [44] or CPHD [44] filters are examples of this. Although these take into account the problems that arise in multiple object estimation, they also suffer from several shortcomings – random finite set approaches discard track identities, or try to propagate them in inefficient ways. Classical approaches are heuristic based and it is not possible to theoretically verify that their population management techniques are correct. The advantages and shortcomings of these methods will be described in section 2.5. The stochastic populations framework [14, 37] has been recently proposed to both maintain track identities while remaining a theoretically principled method, and it will be described in detail later in the chapter.

The remainder of this chapter describes the recursive Bayesian state estimation framework, which will be exploited in subsequent chapters to derive single- and multi-object filters for space situational awareness.

2.2 Bayesian filtering

Filtering is the process through which the probabilistic estimate, or filtering distribution, of the object state is maintained as time passes, by using the dynamical model of the object; and corrected when data is acquired, by using a model of how the sensor observes the object [5]. These models also take into account the uncertainties induced by the sources mentioned above. If these models are available, then the filtering distribution can be obtained by applying the following recursion:

$$p_{k|k-1}(\boldsymbol{x}|\boldsymbol{z}_{1:k-1}) = \int f_k(\boldsymbol{x}|\boldsymbol{x}')p_{k-1}(\boldsymbol{x}'|\boldsymbol{z}_{1:k-1})\,\mathrm{d}\boldsymbol{x}' \tag{2.1}$$

$$p_k(\boldsymbol{x}|\boldsymbol{z}_{1:k}) = \frac{g_k(\boldsymbol{z}_k|\boldsymbol{x})p_{k|k-1}(\boldsymbol{x}|\boldsymbol{z}_{1:k-1})}{\int g_k(\boldsymbol{z}_k|\boldsymbol{x}')p_{k|k-1}(\boldsymbol{x}'|\boldsymbol{z}_{1:k-1})\,\mathrm{d}\boldsymbol{x}'}, \tag{2.2}$$

where $p_{k|k-1}(\boldsymbol{x}|\boldsymbol{z}_{1:k-1})$ is the predicted density at time k; $f_k(\boldsymbol{x}|\boldsymbol{x}')$ is the nonlinear state transition kernel of the system, i.e., the probability of the target being in state $\boldsymbol{x}$ given previous state $\boldsymbol{x}'$; $p_k(\boldsymbol{x}|\boldsymbol{z}_{1:k})$ is the data-corrected or updated density up

to time k; and $g_k(\boldsymbol{z}_k|\boldsymbol{x})$ is the measurement model, or the likelihood of observing measurement $\boldsymbol{z}_k$ conditioned on state $\boldsymbol{x}$. The explicit conditioning on the past measurements will be dropped from here onwards for reasons of succinctness. Equation (2.1) is called the Chapman-Kolmogorov equation, and uses the knowledge about how the process evolves in time to predict the filtering distribution before receiving any additional data. Equation (2.2) is an application of Bayes' rule, and uses the measurement model to integrate the information of any acquired measurements into the filtering distribution. It is clear that further to this, a *prior* distribution $p_0(\boldsymbol{x})$ is required in order to recursively compute the subsequent distributions. This distribution represents any available knowledge about the object state before starting the filtering process, and proper modeling of this initial distribution is essential to obtaining accurate results. All together, this recursion is called the Bayes filter [80], and does not constrain the form of the estimated distributions or the used models.

The evolution of the process of interest through time is considered through the dynamical model. The function $f_k(\boldsymbol{x}|\boldsymbol{x}')$ summarizes the knowledge of how the state of the target evolves through time, and models also any uncertainty on this evolution. Dynamical models range from Brownian motion, where the only source of movement is random; passing by constant velocity or constant acceleration models, where it is assumed that these vector quantities only vary due to unmodeled sources; to sophisticated models for maneuvering targets. A survey of commonly used dynamical models can be found in [75], emphasizing those used for maneuvering targets where it is critical to properly model their motion. In certain cases, the dynamics of the objects that are modeled are very well known. One such case is that of orbital objects, in which case the dynamical models can be borrowed from the physical models describing their motion [7]. This case is studied further in Chapter 3.

The measurement model $g_k(\boldsymbol{z}_k|\boldsymbol{x})$ describes the type of measurements that are acquired by the sensor, conditioned on the object state and possibly other measured properties such as its attitude or reflectivity, and the sensor's own measurement capabilities at the time of the observation. Although the measurements will be used to refine the estimate of the state of the object, it is possible that the sensor can only observe part of the state of interest such that it is not possible to fully resolve it using a single measurement. Additionally, the measurement model takes into account its noise characteristics, in order to incorporate this source of uncertainty into the filtering process. Measurement models can be anything from a fully observed process, to complex non-linear interactions between the observed process and the sensor. In further chapters, measurement models for different filtering problems are

presented, including a way to model the sensor induced uncertainty.

2.3 Single object state estimation

Although (2.1) and (2.2) describe the statistically optimal filter for arbitrary forms of the filtering distributions, in practice computing the integrals becomes intractable unless the functional forms of the probability distributions is constrained. These equations do not have closed form solutions in most cases, a notable exception being that of Gaussian functions. In this section, the most commonly used approaches to tractably use the Bayes filter are presented, starting with the closed form solutions in the Kalman filter family of methods, followed by numerical integration approaches.

The purpose of showing this variety of single object state estimation filters is that not only do they provide the essential building blocks to multiple object state estimation algorithms, but also that according to the application, some filters will be more suitable than others. In further chapters, new filters are derived that are based on the ones described below, which makes it important to introduce them here.

2.3.1 Closed form solutions

The Kalman filter is one of the most widely used solutions to the tractable implementation of the Bayes filter [47]. It propagates the mean and covariance of the distribution of the observed process under certain conditions; namely, that the measurement and dynamical models are linear and their respective random terms are zero mean uncorrelated Gaussian random variables:

$$f_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = F\boldsymbol{x}_{k-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\cdot; 0, Q_k); \tag{2.3}$$

$$g_k(\boldsymbol{z}_k|\boldsymbol{x}_k) = H\boldsymbol{x}_k + \boldsymbol{\nu}_k, \quad \boldsymbol{\nu}_k \sim \mathcal{N}(\cdot; 0, R_k), \tag{2.4}$$

where

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{d}{2}}|\Sigma|^{-\frac{1}{2}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{2.5}$$

is a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$, covariance Σ, evaluated at d-dimensional point $\boldsymbol{x}$; and F and H are the matrices dictating the linear transformations of the dynamical and measurement processes, respectively. If this is the case, the Kalman filter uses the prior mean and covariance of the process $\boldsymbol{\mu}_{k-1}$ and

P_{k-1} to obtain in first instance the predicted mean and covariance

$$\boldsymbol{\mu}_{k|k-1} = F\boldsymbol{\mu}_{k-1} \tag{2.6}$$

$$P_{k|k-1} = FP_{k-1}F' + Q_k. \tag{2.7}$$

From here, when a measurement is obtained, it becomes possible to compute the innovation mean $\boldsymbol{z}_k - H\boldsymbol{\mu}_{k|k-1}$ and its covariance $S = HP_{k|k-1}H' + R_k$, which describe the distribution of the difference between the observed measurement with respect to the expected measurement. This permits the computation of the and the *Kalman gain* $K = P_{k|k-1}H'S^{-1}$, which minimizes the variance of the estimator of the updated mean and covariance

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z}_k - H\boldsymbol{\mu}_{k|k-1}) \tag{2.8}$$

$$P_k = (I - KH)P_{k|k-1}. \tag{2.9}$$

Kalman's original article on this filter tackles the problem from a signal processing point of view, but it is also interesting to consider the problem from a Bayesian statistics point of view, as analyzed by Ho and Lee in [36]. In here, it is shown that if the prior distribution is Gaussian, not only can these statistics be obtained but the complete form of the distribution can be analytically determined to be Gaussian with the parameters shown above. This is because under a Gaussian likelihood, Gaussian functions are conjugate priors with themselves [69].

The main advantages of the Kalman filter are that it is not only robust and principled, but also readily implementable and computationally efficient. However, the requirement that the measurement and dynamical models be linear turns out to be too restrictive for a wide class of problems, which turn out to include space situational awareness, as both the dynamical and measurement models are non-linear [8].

The attractive properties of the Kalman filter, coupled with the urgent need to filter nonlinear systems that was spurred by the need to localize the Apollo spacecraft as it made its way to the moon [30] led to the development of the Extended Kalman filter (EKF) (See e.g. [5]). Rather than requiring linear transformations represented by matrices, general dynamical and measurement models are used:

$$f_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = f(\boldsymbol{x}_{k-1}) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\cdot; 0, Q_k); \tag{2.10}$$

$$g_k(\boldsymbol{z}_k|\boldsymbol{x}_k) = h(\boldsymbol{x}_k) + \boldsymbol{\nu}_k, \quad \boldsymbol{\nu}_k \sim \mathcal{N}(\cdot; 0, R_k). \tag{2.11}$$

Here, the functions f and g are required to be differentiable as the extended Kalman filter relies on the linearization of these models obtained from their first-order Taylor expansions. It can be noted here that more general functions of the form $f(\boldsymbol{x}_{k-1}, \boldsymbol{\epsilon}_t)$ and $h(\boldsymbol{x}_k, \boldsymbol{\nu}_k)$ can be used, but only the simpler case with additive noise is illustrated here for simplicity. The function f is linearized with respect to its parameter to obtain the matrix

$$F_k = \left.\frac{\partial f}{\partial \boldsymbol{x}_{k-1}}\right|_{\boldsymbol{x}_{k-1}=\boldsymbol{\mu}_{k-1}}. \tag{2.12}$$

The mean of the predicted distribution is computed by applying the full state transition function to the prior mean, but the covariance is obtained using the linearized function:

$$\boldsymbol{\mu}_{k|k-1} = f(\boldsymbol{\mu}_{k-1}) \tag{2.13}$$

$$P_{k|k-1} = F_k P_{k-1} F_k' + Q_k. \tag{2.14}$$

Having obtained this, the measurement model is also linearized around the predicted mean to obtain

$$H_k = \left.\frac{\partial g}{\partial \boldsymbol{x}_k}\right|_{\boldsymbol{x}_k=\boldsymbol{\mu}_{k|k-1}}. \tag{2.15}$$

Similarly, the innovation $\boldsymbol{z}_k - h(\boldsymbol{\mu}_{k|k-1})$ is computed with the full nonlinear function, while the innovation covariance and Kalman gain are calculated with the linear approximation: $S_k = H_k P_{k|k-1} H_k' + R_k$ and $K_k = P_{k|k-1} H_k' S_k^{-1}$. This is sufficient to obtain the updated mean and covariance of the distribution:

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z}_k - g(\boldsymbol{\mu}_{k|k-1})) \tag{2.16}$$

$$P_k = (I - KH) P_{k|k-1}. \tag{2.17}$$

It must be stressed that while under the assumptions outlined above the Kalman filter yields distributions that are statistically optimal, the linearization in the EKF causes the resulting distributions to be only approximate. The degree to which the models can be linearized will determine how accurate the obtained filtering distributions will be.

A more recent development in Kalman-like filters is the Unscented Kalman Filter (UKF) [46]. The key idea of this method is that rather than approximating the functions that compose the dynamical and measurement models, it is simpler and more effective to approximate the distribution using a fixed number of samples. The filter proceeds by decomposing the prior into a set of *sigma points* and associated

weights, in such a way that the resulting empirical distribution will have the same statistics as the original distribution; then propagating these points through the full nonlinear functions; and finally using the resulting points to compute the statistics required to obtain the filtering distribution. This approach differs from particle filtering techniques in that the set of sigma points is chosen in a deterministic way, and the weights do not indicate probabilities.

To use the prior to obtain a set of sigma points, the state must be extended to include the noise terms in the transition kernel $f_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = f(\boldsymbol{x}_{k-1}, \boldsymbol{\epsilon}_k)$, by extending the mean and covariance of the distribution with those of $\boldsymbol{\epsilon}_k$. A common way to obtain these points is to use the Cholesky decomposition of its covariance matrix to obtain a set of points which have the same statistics $\boldsymbol{\mu}_{k-1}$ and P_{k-1}. If the i-th column of the Cholesky decomposition of P_{k-1} is denoted $\boldsymbol{\sigma}_i$, it is straightforward to verify that the distribution of $2N + 1$ points

$$\begin{aligned} \boldsymbol{x}_{k-1}^{0} &= \boldsymbol{\mu}_{k-1} \\ \boldsymbol{x}_{k-1}^{(i)} &= \boldsymbol{\mu}_{k-1} + \boldsymbol{\sigma}_i, \quad i = 1, 2, \ldots, N \\ \boldsymbol{x}_{k-1}^{(N+i)} &= \boldsymbol{\mu}_{k-1} - \boldsymbol{\sigma}_i, \quad i = 1, 2, \ldots, N, \end{aligned} \tag{2.18}$$

where N is the dimension of the extended state, has the desired mean and covariance. The predicted mean and covariance are then obtained by propagating these sigma points through the transition kernel and computing the statistics:

$$\boldsymbol{\mu}_{k|k-1} = \frac{1}{2N+1} \sum_{i=0}^{2N+1} f(\boldsymbol{x}_{k-1}^{(i)}), \tag{2.19}$$

$$P_{k|k-1} = \frac{1}{2N+1} \sum_{i=0}^{2N} (f(\boldsymbol{x}_{k-1}^{(i)}) - \boldsymbol{\mu}_{k|k-1})(f(\boldsymbol{x}_{k-1}^{(i)} - \boldsymbol{\mu}_{k|k-1})', \tag{2.20}$$

where f acts on the extended state vector instead of the original state vector and the random term. To obtain the updated term, these are extended with the observation noise term $\boldsymbol{\nu}_k$ in the measurement model $g_k(\boldsymbol{z}_k|\boldsymbol{x}_k) = h(\boldsymbol{x}_k, \boldsymbol{\nu}_k)$. Following the process outlined above, the extended covariance is again decomposed to obtain the $2M + 1$ predicted sigma points $\boldsymbol{x}_{k+1|k}^{(i)}$ and the predicted observation $\hat{\boldsymbol{z}}_k$ is obtained:

$$\hat{\boldsymbol{z}}_k = \frac{1}{2M+1} \sum_{i=0}^{2M} h(\boldsymbol{x}_{k+1|k}^{(i)}), \tag{2.21}$$

with M the dimension of the extended state space. The expected observation is then used alongside the sigma points to obtain the innovation covariance S, and the

state-observation cross-correlation P_{xz}:

$$S = \frac{1}{2M+1}\sum_{i=0}^{2M}(h(\boldsymbol{x}_{k+1|k}^{(i)}) - \hat{\boldsymbol{z}}_k) \tag{2.22}$$

$$P_{xz} = \frac{1}{2M+1}\sum_{i=0}^{2M}(\boldsymbol{x}_{k+1|k}^{(i)} - \boldsymbol{\mu}_{k|k-1})(h(\boldsymbol{x}_{k+1|k}^{(i)}) - \hat{\boldsymbol{z}}_k)'. \tag{2.23}$$

From this, the Kalman gain can be computed as

$$K = P_{kz}S^{-1}, \tag{2.24}$$

from which the updated mean and covariance can be obtained:

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z} - \hat{\boldsymbol{z}}_k) \tag{2.25}$$

$$P_k = P_{k|k-1} - KSK'. \tag{2.26}$$

An advantage of using filters in this family have to do with the fact that since the full dynamical and measurement models are used rather than approximations, certain biases can be eliminated. In particular, linearizing the commonly used transformation between polar and Cartesian co-ordinates has been shown to yield biased results in the EKF, which does not happen in the UKF [45]. Additionally, deriving and programming the Jacobian matrices required in the EKF is not necessary, which is an intensive and error-prone process.

If priors which can be reasonably represented with a mean and a covariance can be used, the UKF is an attractive method as it is simple and computationally efficient. However, it can be difficult to apply this method if the distributions that are used cannot be represented like this, as is the case in the priors presented in chapter 3.

Since the methods shown above represent distributions through a mean and a covariance, they cannot appropriately propagate multimodal distributions. Even if a distribution is unimodal, the performance of the filter will suffer when the shape of the distribution does not resemble that of a Gaussian. In order to solve these problems, the Gaussian sum filter was proposed, which is based on the observation that it is possible to use Gaussian mixtures to approximate a wide range of distributions [1, 88]. Gaussian sum filters represent the filtering distribution as a sum of

weighted Gaussian distributions:

$$p_k(\boldsymbol{x}|\boldsymbol{z}_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_k^{(i)}, P_k^{(i)}). \tag{2.27}$$

Prediction is applied to each individual Gaussian term in the same way as the previously mentioned filters, and weights are left unchanged. For update, the update operation of the above-described filters is also applied to each individual Gaussian, after which the weights are updated based on the individual innovation and its covariance:

$$w_{k+1}^{(i)} = \frac{w_k^{(i)} \mathcal{N}(\boldsymbol{z} - h(\boldsymbol{\mu}_{k|k+1}^{(i)}); 0, S_k^{(i)})}{\sum_{j=1}^{N} w_k^{(j)} \mathcal{N}(\boldsymbol{z} - h(\boldsymbol{\mu}_{k|k+1}^{(j)}); 0, S_k^{(j)})}. \tag{2.28}$$

A big advantage of closed-form solutions is that the parameters of the posterior distribution can be found even if the mismatch between the prior distribution and the measurement likelihood always exists. This is in contrast to numerical methods, where Gaussians are essentially truncated after a certain distance from the mean, such that the product of two Gaussians can numerically be zero even if it is not the case theoretically. Unfortunately, this comes with the imprecision that is added by the involved approximations.

2.3.2 Numerical integration

In cases where the state space of the variable to estimate is sufficiently small, the integrals in (2.1) and (2.2) can be solved numerically. If the state space is discrete, these can be calculated for each possible state, and the filter is called the discrete Bayes filter. If it is hybrid or continuous, it is first discretized into bins, and a representative point in each of these bins is used for prediction and update. This method is called the *histogram filter* [94].

Since it is common for some regions of the state space to concentrate lower cumulative probability than others, the state space can be decomposed unevenly to represent the regions with higher likelihood with greater granularity, while using a more compact representation for regions that don't accumulate a lot of probability. For this, methods such as quad- or oct-trees can be used [94]. Another interesting method is that of optimal stochastic quantization, which learns a discrete representation of the state space which has higher resolution in the higher likelihood regions [4].

The main disadvantage of solving Bayes' filter numerically is that as the volume and dimension of the state space increases, the problem becomes increasingly un-

tractable from the computational point of view. This issue, usually called the *curse of dimensionality*, has motivated the development of Monte Carlo methods which are more tractable in high-dimensional state spaces.

2.3.3 Monte Carlo methods

An alternative to parametric filters such as the ones described above is the family of Monte Carlo methods. These algorithms rely on solving the intractable integrals which determine the predicted and updated state distributions using Monte Carlo integration, which consists on making use of a weighted sample representation of the probability distribution of the object state, and then using the weighted samples to approximate the continuous integral as a discrete sum. This is useful since if it is possible to draw N samples $\{\boldsymbol{x}_k^{(i)}\}_{i=1}^N$ from a distribution of interest, it is possible to estimate expected values using the following approximation [18]:

$$\mathrm{E}_{p_k}(f) = \int f(\boldsymbol{x}) p_k(\boldsymbol{x}) \mathrm{d}\,\boldsymbol{x} \approx \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_k^{(i)}), \tag{2.29}$$

where $\mathrm{E}_{p_k}(f)$ denotes the expected value of function f under probability distribution p_k.

Representing a distribution with samples is not only useful to compute expected values (from where the statistics of the distribution can be obtained), but also to get an idea of the shape of the distribution, as areas with higher concentration of particles integrate to higher probability. Equation (2.29) assumes that it is possible to obtain independent, identically distributed (IID) samples from the probability distribution p_k. Generally, however, obtaining samples from arbitrary distributions is not straightforward. The importance sampling framework is commonly used to overcome this difficulty. It is based on the principle that the above expectation is equivalent to

$$\mathrm{E}_{p_k}(f) = \int \frac{p_k(\boldsymbol{x})}{\pi(\boldsymbol{x})} f(\boldsymbol{x}) \pi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2.30}$$

where π, called the importance sampling function, is a probability distribution which can be easily sampled from, with support overlapping that of p_k. This suggests that the expectation can be computed by sampling N particles from π and then using

equation (2.29) to obtain

$$\mathrm{E}_{p_k}(f) \approx \sum_{i=1}^{N} w_k^{(i)} f(\boldsymbol{x}_k^{(i)}), \quad \text{with} \tag{2.31}$$

$$w_k^{(i)} = \frac{p_k(\boldsymbol{x}_k^{(i)})/\pi(\boldsymbol{x}_k^{(i)})}{\sum_{j=1}^{N} p_k(\boldsymbol{x}_k^{(j)})/\pi(\boldsymbol{x}_k^{(j)})}. \tag{2.32}$$

An alternative way of seeing this is that the probability distribution is being approximated with a set of *weighted samples* $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$, with higher weights representing higher likelihood for each particular sample. As the importance sampling function is used to obtain samples to represent the updated distribution, the filter efficiency will improve if these two distributions are close. If it is possible to directly sample from the updated distribution (2.2), e.g., if $p_{k-1|k-1}(\boldsymbol{x})$ is Gaussian and $g_k(\boldsymbol{z}_k|\boldsymbol{x})$ and $f(\boldsymbol{x}|\boldsymbol{x}')$ are linear and Gaussian and methods such as the ones described in section 2.3.1 are used, then the proposal is said to be optimal as it minimizes the variance of the importance weights [18].

To maintain a representative sample of the distribution, resampling is usually performed which replaces low weighted particles by particles in areas of higher likelihood, thus increasing the resolution of the distribution in the regions where more precision is required. This is done by replacing the weighted sample $\{\tilde{\boldsymbol{x}}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ by an equally weighted sample $\{\boldsymbol{x}_k^{(i)}, \frac{1}{N}\}_{i=1}^N$, with probability $\Pr(\boldsymbol{x}_k^{(i)} = \tilde{\boldsymbol{x}}_k^{(j)}) = w_k^{(j)}$. These two sets of particles approximate the same original distribution.

In order to obtain samples from the distribution, methods such as those in the Markov chain Monte Carlo (MCMC) family can be used. For example, the Metropolis-Hastings algorithm requires only a conditional proposal distribution q and a function that is proportional to the probability distribution to sample from, $f(x) \propto p(x)$ [34]. The fact that the probability distribution only needs to be known up to a constant of proportionality is useful in this case as it means that the denominator in (2.2) does not need to be computed, for instance. The method approximates the distribution $p(\cdot)$ by starting at an initial random sample x_0 and iteratively sampling from a proposal kernel conditioned on the current point, $\tilde{x}_k \sim q(\cdot|x_{k-1})$. This is accepted as the next sample $x_k = \tilde{x}_k$ if $p(\tilde{x}_k) \geq p(x_{k-1})$. If the probability is lower, then it is accepted with probability $f(\tilde{x}_k)/f(x_{x-1})$ and rejected otherwise, in which case $x_k = x_{k-1}$. Since this method tends to generate autocorrelated chains, and needs to generate a number of samples before it achieves the desired stationary distribution, it requires the generation of a number of samples before converging to it, in addition to the usual requirement of *thinning*, or only taking one sample every

N samples generated from the obtained sequence, in order to avoid these undesired correlations.

The MCMC family of methods rely on exploring regions of the state space with high probabilities, as they need to provide more samples in these regions than in others. Simple Brownian motion can be used as a proposal kernel in a process known as Random Walk Monte Carlo, but the process will be aided by using more information about the target distribution such as its gradient, as this enables the exploration of higher likelihood areas. Methods that exploit this include the Metropolis Adjusted Langevin Algorithms (MALA), which uses Langevin dynamics which make use of the gradient of the logarithm of the posterior to create a stochastic sequence that converges to this distribution [74]. Hamiltonian Monte Carlo, in turn, uses Hamiltonian dynamics to explore the state space [62]. This involves using an auxiliary momentum variable, and has also been shown to perform very well.

In cases where it is simple to sample from sub-groups of variables of the target distribution, and it is possible to compute the conditional distributions of the remainder of the variables on this sub-group, Gibbs sampling may be used [9]. This is an effective way of reducing the dimensionality of the problem, which greatly reduces the complexity of MCMC methods.

More methods exist to obtain samples from the desired distribution, including deterministic surrogate and optimization-based methods. An extensive survey of stochastic simulation methods such as these can be found in [65].

As the time step k in (2.1) and (2.2) increases, generating samples using these methods becomes increasingly onerous as the complete chain of samples must be generated from the beginning up to the current time step. Sequential Monte Carlo methods, however, allow for the computation of the current filtering distribution conditioned on the previous belief, which is ideally suited to the filtering problem as only one sample needs to be generated instead of having to recompute the entire particle trajectory [18]. This filter is initialized with a sample of the prior distribution, $\{\boldsymbol{x}_0^{(i)} \sim p_0(\cdot)\}_{i=1}^N$, with equal weights $w_0^{(i)} = 1/N, i = 1, \cdots, N$. At time step k, when a measurement $\boldsymbol{z}_k$ is received, it obtains the samples

$$\boldsymbol{x}_k^{(i)} \sim f(\cdot|\boldsymbol{x}_{k-1}^{(i)})$$

followed by the importance weights

$$w_k^{(i)} = \frac{\tilde{w}_k^{(i)}}{\sum_{i=1}^N \tilde{w}_k^{(i)}}, \quad \tilde{w}_k^{(i)} = g(\boldsymbol{z}_k|\boldsymbol{x}_k^{(i)}) w_{k-1}^{(i)},$$

which is followed by resampling to avoid particle degeneracy. As it can be seen, the classical bootstrap filter [29] uses the Markov transition kernel for each particle as an importance function, but does not use any information about the received measurement. Proposal distributions which make use of the measurement are called *fully adapted*, and will naturally approximate the desired distribution better as they use all the available information up to the current time step.

A way to use a fully adopted proposal in the particle filtering framework is to use a MCMC rejection method to approximate the optimal distribution when sampling the next particle [41]. The advantage of this approach is that the optimal proposal is approximated directly, minimizing the variance of the particle weights. With this, however, comes a highly increased computational cost. An alternative to this is the Auxiliary Particle Filter [66]. In here, an additional step is performed to randomly select a particle from the previous time step such that samples of higher likelihood are obtained, and integrates the measurement likelihood into the proposal distribution. This additional step yields more particles in the more informative regions of the state space. After this, the proposal weights are computed as before. This filter has been shown to strike a balance between computational efficiency and filter performance.

The issues involved with using particle filters in SSA have to do with the fact that samples are needed from the Bayes filter recursion. The transition kernel tends to be very narrow, as very little noise is needed in dynamical models for orbiting objects. This is usually problematic in Monte Carlo approaches and tends to be solved by using *bridging densities* [18]. These are intermediate densities which converge at a given rate to the desired density. The problem with this is that the computational cost of the resulting algorithm is greatly increased due to the intermediate sampling steps.

2.4 Case study: A filter for laser ranging

In this section, single object filtering is illustrated by applying it to data acquired from a ground-based laser ranging station that measures the distance from it to satellites equipped with retroreflectors. The output of the range-only laser sensor at the Herstmonceux Space Geodesy Facility is analysed in order to design a sensor

Section 2.4 uses material from 'Likelihood modelling of the Space Geodesy Facility laser ranging sensor for Bayesian filtering' [86], C. Simpson, A. Hunter, S. Vorgul, E. Delande, J. Franco, D. Clark, published in the proceedings of the Sensor Signal Processing for Defence (SSPD) conference, used with permission.

model for filtering purposes. The sensor model is then exploited for the design of a single-target Bayesian filter, comparing a Kalman filter and a particle filter.

The Space Geodesy Facility in Herstmonceux (East Sussex, UK) is a multi-technique geodetic observatory operating an SLR station, an absolute gravimeter and several Global Navigation Satellite System (GNSS) receivers. Along with forty other similar sites around the world, the SGF in Herstmonceux forms part of the International Laser Ranging Service (ILRS) [64]. The SLR technique, used primarily for geodetic purposes, measures the time of flight of short laser pulses as they travel between the observing stations and orbiting satellites equipped with retroreflectors [10, 83]. Satellites routinely tracked by the ILRS network include low Earth orbiters with scientific payloads (e.g. Grace, Jason-3, Swarm), passive geodetic targets (e.g. LAGEOS, LARES), and various GNSS constellations (e.g. GLONASS, BeiDou, GPS). Capable of providing measurements with sub-centimeter accuracy and precision, SLR is one of the four space geodetic techniques contributing to the realization of the International Terrestrial Reference Frame [2]. Beyond geodetic applications, SLR can also be employed to track uncooperative space debris objects (i.e. no retroreflectors present) [50, 102].

An Nd:Van pulsed laser (1 KHz repetition rate, 10 ps FWHM pulse width, 1.1 mJ/pulse) at the frequency-doubled wavelength of 532 nm is employed at the SGF laser station. The receiver telescope is a 0.5 m Cassegrain reflector equipped with a Single Photon Avalanche Diode (SPAD) detector. The timing measurements are provided by a home built event timer of 1 ps resolution and 5 ps precision. A strictly single-photon tracking policy is followed at SGF for all satellite targets, whereby the energy levels of the returned pulses are controlled and limited to ensure that, on average, only a single photon is contained in each reflected pulse. This ensures that the laser retroreflector arrays carried onboard the satellite targets are sampled in their entirety, with no preferential detections obtained from points closer to the ground station. In order to limit the negative impact of background and dark noise events, the detector is gated shortly earlier (typically 100 ns) than the predicted range to the satellite. This is necessary due to the high sensitivity of the sensor and the present background radiation. The distribution of returns, excluding actual satellite reflections, are adequately described with a negative exponential distribution, as the detection events follow Poisson statistics [83]. The specific characteristics of the distribution of detected pulses from the satellite targets depend on the shape and orientation of the laser retroreflector arrays.

Three datasets collected from the SGF laser, named 746, 540, 737 for different satellite passes, can be seen in Figure 2.1. The identities of the observed satellites

is known, and the ground truth, shown in the figure is obtained from an available catalog. These figures illustrate the typical features of the raw ranging data collected at SGF, though they all present a very noticeable skewness in the data distribution around the ground truth. The data residuals are depicted in Figure 2.2. In particular, batch 737 has an atypical shape in the lower range values due to a temporal problem in the receiver hardware caused by laser overlap, which happens when a pulse is fired at the same time a detector is gated. The pulse backscatters off the atmosphere and triggers the detector. This run was recorded when the overlap avoidance routine was disabled.

2.4.1 Filter design

A simple constant velocity motion model will be used to track the position and velocity of each object. The state will be denoted $\boldsymbol{x}_t = (r, \dot{r})$, and its dynamics are modeled as

$$\boldsymbol{x}_t = \begin{bmatrix} 1 & \Delta_t \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \boldsymbol{n}_t, \tag{2.33}$$

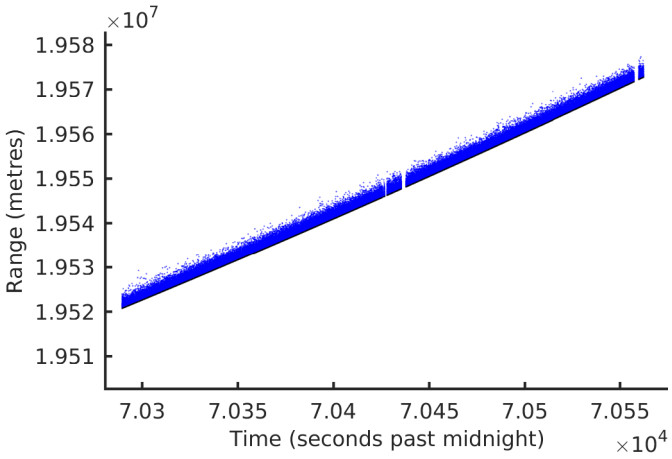
with $\boldsymbol{n}_t \sim \mathcal{N}(\mathbf{0}, Q_t)$ is the process noise vector. The measurement model is more interesting, as it is evident from the residuals in Figure 2.2 that the noise distribution does not have a Gaussian form. In order to obtain a suitable likelihood function, an exponential curve was fit to the residuals of batch 746, obtaining the following estimated relation:

$$\ell(z_t|\boldsymbol{x}_t) \propto \exp(-2.811 \times 10^{-4}(z_t - r_t)). \tag{2.34}$$

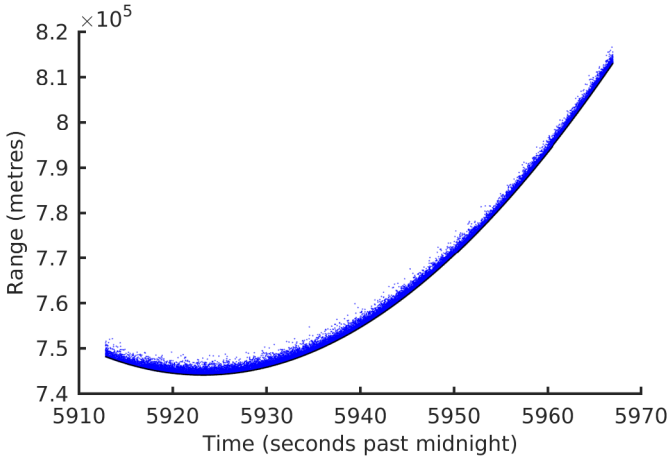
The resulting curve can be seen in Figure 2.3.

In order to implement a Kalman filter, a Gaussian distribution was fit to the residuals of the same batch in order to produce a linear observation model. The results of this filter can be seen in Figure 2.4. From this figure, it can be seen that since the filter expects a symmetrical distribution for the measurement noise, it produces biased estimates.

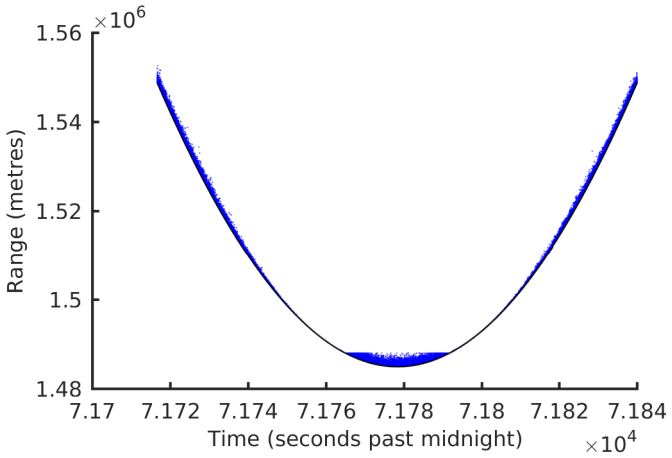
A simple bootstrap filter was also implemented, as described in Section 2.3.3. The filter was applied to datasets 540 and 737, using the likelihood function estimated from dataset 746. As it can be seen in Figure 2.1, batch 540 has a very similar noise distribution to the training set while batch 737 has some artifacts resulting in a more complicated noise structure. The results of applying the filter with dataset 540 can be seen in Figure 2.5, where it can be seen how the effects of the noisy



(a) Batch 746

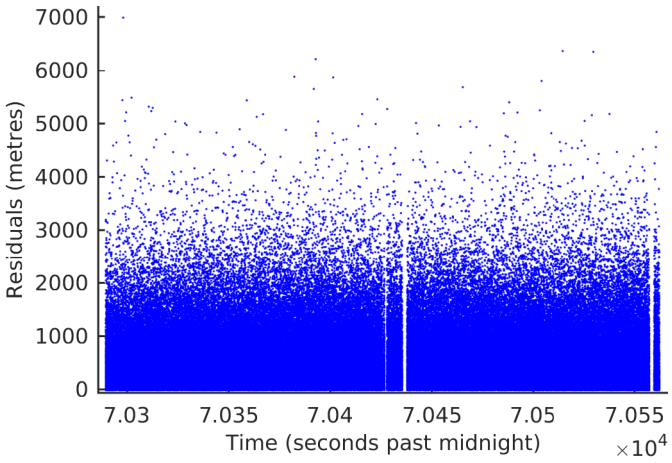


(b) Batch 540

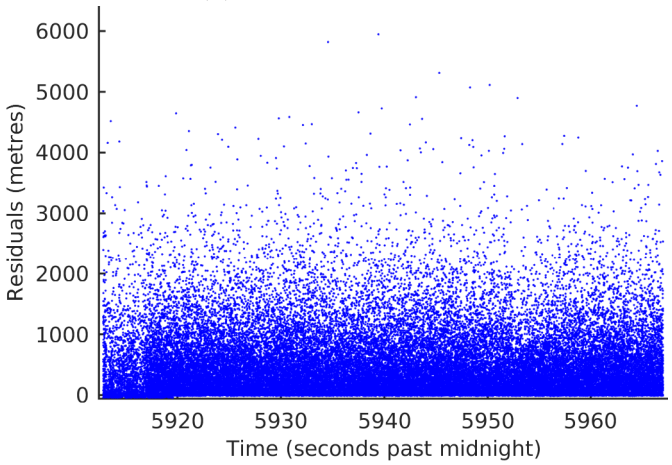


(c) Batch 737

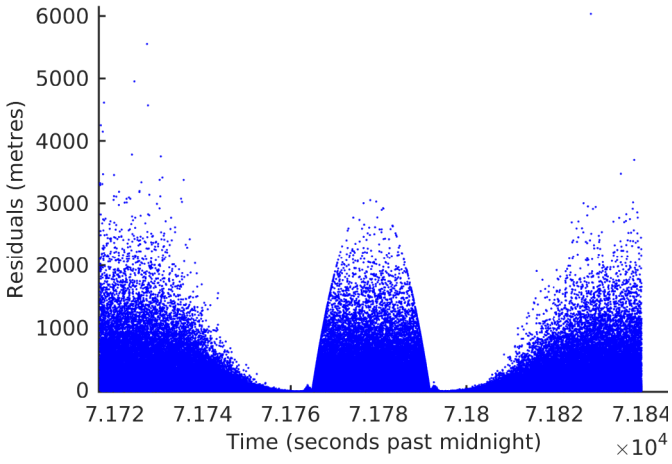
Figure 2.1: SLR measurements. Ground truth (black), measurements (blue).



(a) Batch 746



(b) Batch 540



(c) Batch 737

Figure 2.2: SLR residual data. Measurements in blue.

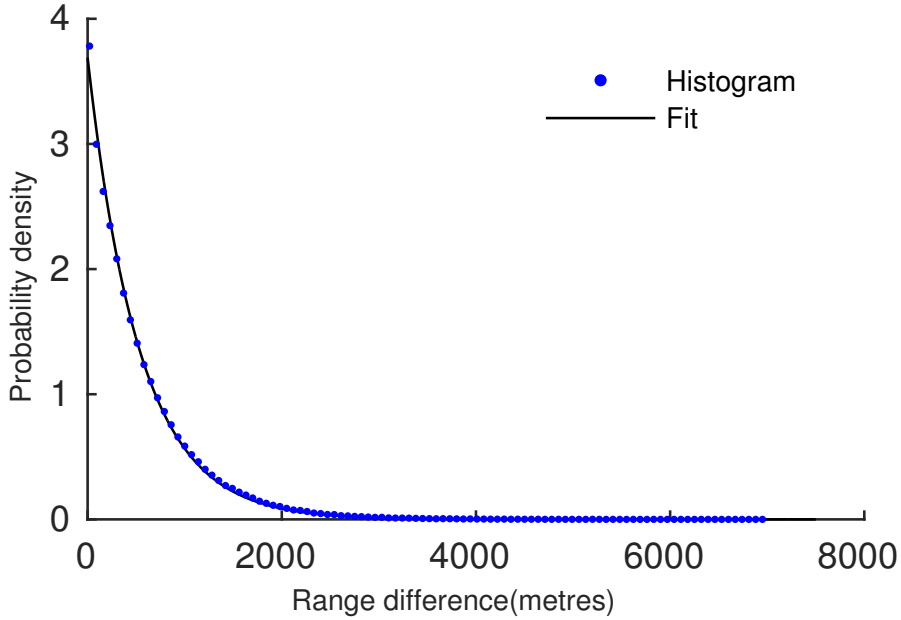


Figure 2.3: Estimated measurement likelihood

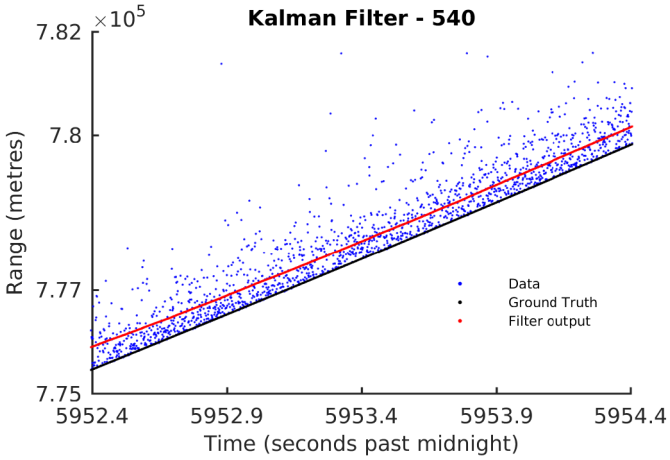


Figure 2.4: Results of Kalman filtering on batch 540

measurements are reasonably filtered out. Figure 2.6 shows the results of the filter on batch 737. Here, it can be seen that the filter is also robust to sudden changes in the distribution of the noise.

This sample filtering application shows how it is possible to handle different noise profiles by adequately modeling the measurement likelihood functions, and how Kalman filtering is not always applicable. The results indicate that the estimates are consistent with the object states found in the catalog, in spite of the challenging noise profile.

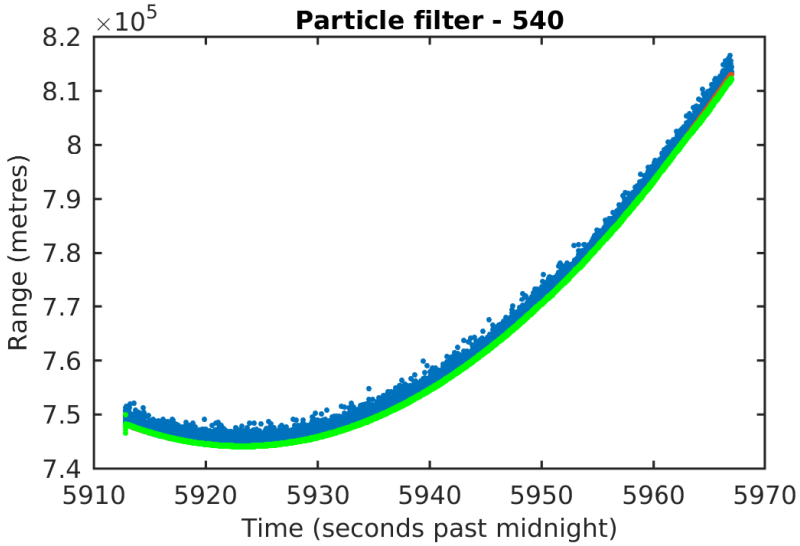
2.5 Multiple object state estimation

Multiple object estimation, or multi-target tracking, is the problem of simultaneously estimating the state of a group of objects of interest as it evolves through time, and its unknown and time-varying size. This is particularly interesting as it is general enough to study multiple problems of interest, including disciplines such as simultaneous localization and mapping (robotics) [53], biological microscopy [25], and defense [58]. Estimating the state of such a system, however, is rarely as easy as estimating the states of each one of its components, due to a multiple of problems which include the uncertainty of track-to-measurement associations, the possible presence of spurious measurements in the data which are not produced by any object of interest, and the possibility of missed detections, all of which increase the difficulty of the estimation process [5].

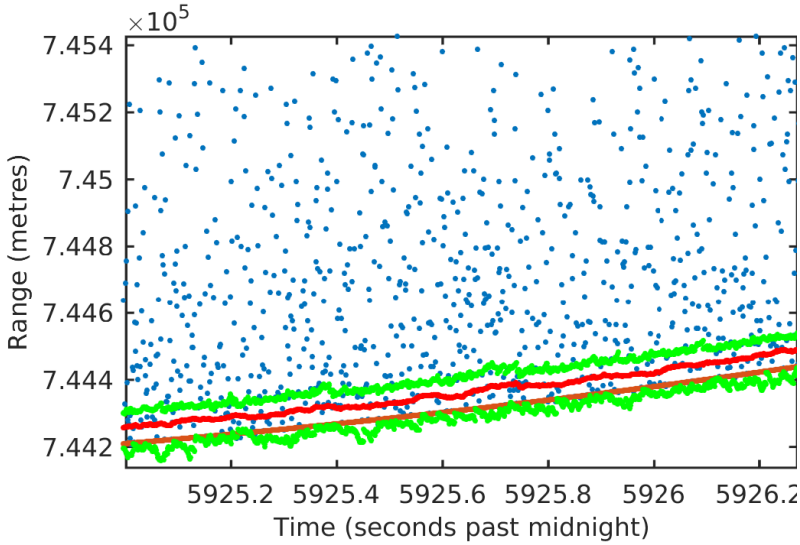
Several approaches to solve the multi-object estimation problem exist, which can be broadly divided into three categories. The first category can be referred to as the 'classical' category, and attempts to use heuristics based on data association techniques to assign measurements to single-object filters. The second category is made up of filters based on Finite Set Statistics (FISST), which attempt to track the whole population by estimating probability densities defined on *random sets* rather than *random vectors*. The third category is based on a new formulation based on the concept of distinguishability in stochastic processes. These techniques will be described below.

2.5.1 Classical approaches

Classical methods of multi-target tracking are based on heuristic systems that manage a group of single target filters, assigning them measurements according to data association heuristics. The Joint Probabilistic Data Association Filter (JPDAF)

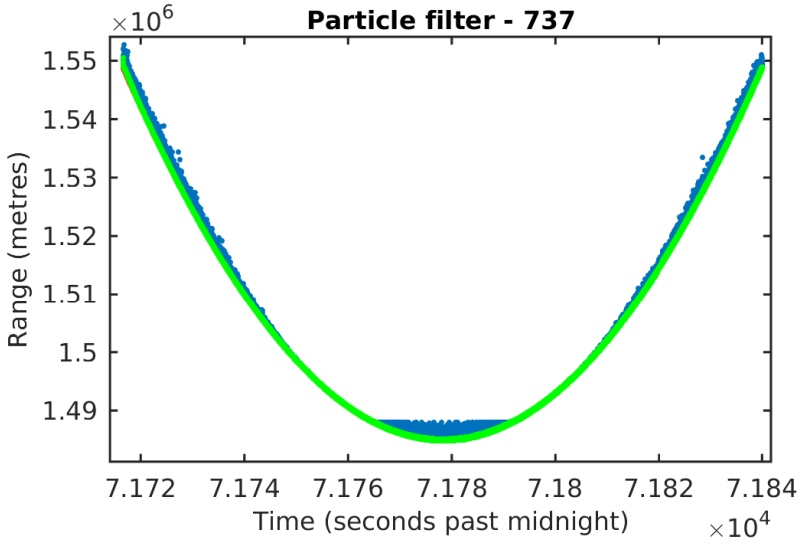


(a) Maxima and minima of the filtering distribution throughout the estimation (green), measurements (blue)

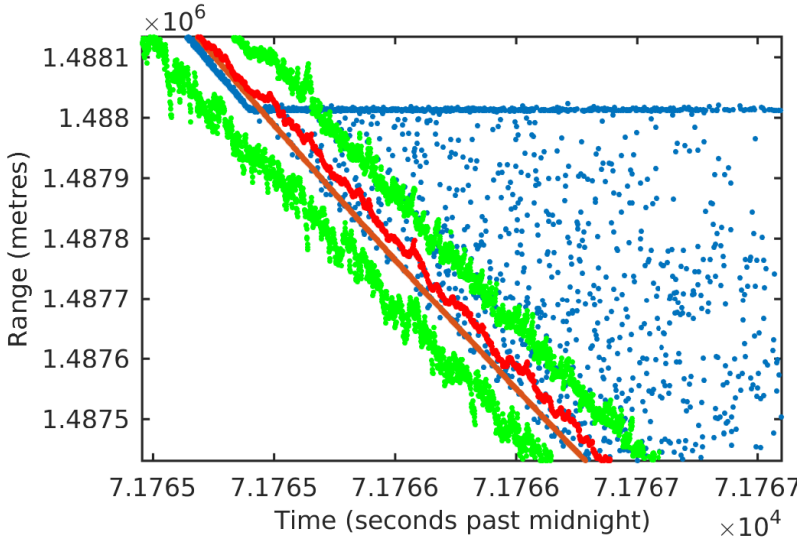


(b) Detail of filtering results. Measurements and extrema as above, ground truth (orange), estimate (red).

Figure 2.5: Results of the particle filter on pass 540



(a) Maxima and minima of the filtering distribution throughout the estimation (green), measurements (blue)



(b) Detail of filtering results. Measurements and extrema as above, ground truth (orange), estimate (red).

Figure 2.6: Results of the particle filter on pass 737

assumes that the number of targets is known, and assign measurements as either being produced by a particular track or being a false alarm, leaving the remainder of the tracks as tracks with a missed detection [5]. The data association mechanism is based on the Mahalanobis distance between tracks and measurements. The clear disadvantage of this technique is that the number of tracks must be known beforehand by the operator, and is assumed to remain constant during the estimation process - In many applications, target appearance and disappearance are essential components of the dynamics of the multi-object system.

The Multiple Hypothesis Tracking (MHT) filter [71] is one of the most commonly used target tracking filters today, perhaps due to the fact that it is a rather straightforward extension of single-target filtering techniques. It incorporates target appearance and disappearance into the filtering process by considering, for each measurement, whether it was originated by either one of the previous tracks, a false alarm, or a new track.

Each group of associations where each measurement is assigned to one of these categories is called a hypothesis, and single target filters are run to evaluate the new multi-target state per hypothesis. The likelihood of each hypothesis is then evaluated by combining the individual filter likelihoods, and taking into account the likelihood of the hypothesized false alarms and misdetections. Further heuristics help curb the geometric growth of hypotheses, which would make the algorithm prohibitively expensive to run after some time. Although this approach considers the necessary components to make a useful multi-target tracker, the extensive use of heuristics adds to the problem the need of tuning a number of additional parameters and makes it hard to validate the filter analytically (e.g., by providing convergence results).

It is worth noting here that different domain-specific methods exist for tracking which have not been mentioned here. For instance, active contour methods such as the one described in [48] can be used to effectively track moving objects in images based on edge motion, or acoustic trackers such as those based on iterative time reversal techniques are used to interactively focus energy on targets of interest with active sonar systems [68]. The focus of this thesis, however, is to analyze general tracking methodologies rather than these domain-specific methods.

2.5.2 Random finite set solutions

Finite Set Statistics (FISST) is an approach which generalizes the single-object Bayes filter to multiple targets by using random finite sets (RFS) rather than ran-

dom vectors. It introduces concepts such as the set integral and probability generating functionals which permit the direct manipulation of multi-object distributions without the need to use heuristics and data association [58]. The first practical filter produced with this framework is the Probability Hypothesis Density (PHD) Filter [56], which propagates the first order moment of the multi-target distribution to estimate both the number of objects of interest in the scene and their individual states. The resulting filter is not computationally burdensome and the resulting expressions for prediction and update are intuitive, which simplifies its implementation and the interpretation of its results. A downfall of the PHD filter, however, is its strong assumption that the prior distribution is a Poisson multi-target process, which means that all objects are Independent Identically Distributed (IID) and that the cardinality distribution is Poisson. Poisson probability mass functions have means equal to their variances, which causes the estimate of the number of targets to be somewhat unstable when object appearance and disappearance are frequent. Recent developments have attempted to overcome this limitation by using more general distributions such as the negative binomial distribution [81].

The Probability Hypothesis Density (PHD) Filter [56] propagates the first moment $D(x)$ of the multi-target posterior, known as the PHD or intensity, which is a function defined on the single-target state space which indicates the expected number of targets in any of its regions. If the multitarget probability distribution is denoted $p_k(\boldsymbol{X})$, then the PHD satisfies

$$\int_S D(x)\ \mathrm{d}x = \int |\boldsymbol{X} \cap S| p_k(\boldsymbol{X})\ \delta\boldsymbol{X} = N_k(S), \tag{2.35}$$

where $N_k(S)$ denotes the expected number of targets in set S and $|S|$ is the cardinality of S. The integral of the form $\int \ldots \delta\boldsymbol{X}$ is a *set integral* [56]. The process and measurement models used by the PHD Filter are based on the following assumptions:

1. From time step $k-1$ to time step k, each target $\boldsymbol{x}_{k-1}$ survives with probability $P_S(\boldsymbol{x}_{k-1})$ (the *probability of survival*), evolving into $\boldsymbol{x}_k \sim f_k(\cdot|\boldsymbol{x}_{k-1})$ or disappears with probability $1 - P_S(\boldsymbol{x}_{k-1})$.

2. New targets may appear at each time step according to an independent process.

3. Each target $\boldsymbol{x}$ produces a measurement $\boldsymbol{z} \sim g_k(\cdot|\boldsymbol{x})$ with probability $P_D(\boldsymbol{x})$ (the *probability of detection*) or is not detected with probability $1 - P_D(\boldsymbol{x})$.

4. False alarms are produced at each time step according to a certain clutter distribution.

These assumptions are synthesized in the following RFS process and measurement models:

$$\boldsymbol{X}_k = \boldsymbol{\Gamma}_k \cup \bigcup_{\boldsymbol{x} \in \boldsymbol{X}_{k-1}} \boldsymbol{Y}_{k|k-1}(\boldsymbol{x}) \tag{2.36}$$

$$\boldsymbol{Z}_k = \boldsymbol{K}_k \cup \bigcup_{\boldsymbol{x} \in \boldsymbol{X}_k} \boldsymbol{\Theta}_k(\boldsymbol{x}) \tag{2.37}$$

Where $\boldsymbol{\Gamma}_k$ and $\boldsymbol{K}_k$ are the birth and clutter random finite sets, respectively, and

$$\boldsymbol{Y}_{k|k-1}(\boldsymbol{x}) = \begin{cases} \{\boldsymbol{x}'\}, & \text{with probability } P_S(\boldsymbol{x}) \text{ and } \boldsymbol{x}' \sim f_k(\cdot|\boldsymbol{x}) \\ \emptyset, & \text{with probability } 1 - P_S(\boldsymbol{x}) \end{cases} \tag{2.38}$$

$$\boldsymbol{\Theta}_k(\boldsymbol{x}) = \begin{cases} \{\boldsymbol{z}\}, & \text{with probability } P_D(\boldsymbol{x}) \text{ and } \boldsymbol{z} \sim g_k(\cdot|\boldsymbol{x}) \\ \emptyset, & \text{with probability } 1 - P_D(\boldsymbol{x}). \end{cases} \tag{2.39}$$

A derivation of the PHD filter starting from these assumptions can be found in [56]. The resulting prediction and update equations are the following:

$$D_{k|k-1}(\boldsymbol{x}) = \gamma(\boldsymbol{x}) + \int P_S(\boldsymbol{x}') f_{k-1}(\boldsymbol{x}|\boldsymbol{x}') D_{k-1}(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}' \tag{2.40}$$

$$D_k(\boldsymbol{x}) = (1 - P_D(\boldsymbol{x})) D_{k|k-1}(\boldsymbol{x}) + \sum_{\boldsymbol{z} \in \boldsymbol{Z}_k} \frac{P_D(\boldsymbol{x}) g_k(\boldsymbol{z}|\boldsymbol{x}) D_{k|k-1}(\boldsymbol{x})}{c(\boldsymbol{z}) + \int g_k(\boldsymbol{z}|\boldsymbol{x}') D_{k|k-1}(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x}'}, \tag{2.41}$$

Here, $\gamma(\boldsymbol{x})$ is the PHD of the birth process and $c(\boldsymbol{z})$ is the PHD of the clutter process. In many applications, priors for newborn targets are not suitable to adequately describe appearing targets. In these cases, an alternative strategy can be used with measurement driven births [40]. In this approach, when no *a priori* information is available on where targets are likely to appear, measurements are used to determine likely positions new target appearances.

As in the case of the single target Bayes filter, an appropriate form for $D(\boldsymbol{x})$ must be chosen in order to use equations (2.40) and (2.41) to implement a tractable filter. The most common approaches to do this are Gaussian mixture implementations [97] and SMC implementations [98], mirroring the single target tracking case.

In addition to estimating the state of populations of objects, the PHD filter has been extended to estimate more complex phenomena such as groups of objects with correlated motion (*group targets*) [90] and targets which can generate multiple measurements (*extended targets*) [89] by modeling them as independent spatial

cluster point processes. As a particular case of these filters, the single-cluster PHD filter was developed in order to estimate populations of objects conditioned on a single-object random variable, which is also unknown [91]. This lends itself well to problems where the population of objects is observed through a sensor which has unknown state, as it is necessary to estimate the state of the sensor in order to produce unbiased estimates of the observed population [73].

The single-cluster PHD filter has been used in many interesting applications. Ristic et al have used it in order to calibrate sensors using non-cooperative targets [73], while Lee et al. have applied to the robotics problem of simultaneous localization and mapping [52, 53, 54]. Schlangen et al have used it to solve the problem of estimating the position of intra-cellular proteins observed using fluorescence microscopy, while simultaneously correcting for microscope drift [25, 82]. Hagen et al. applied it to image plane tracking of objects observed from a telescope, while simultaneously stabilizing it [33], and Houssineau et al. applied it to simultaneous 3D tracking and camera calibration from video data.

An important limitation of the PHD filter is that it only propagates the first moment of the multi-object posterior. Due to this, several attempts have been made to create more informative filters using the FISST framework by propagating more information. The Cardinalized Probability Hypothesis Density (CPHD) filter, for instance, propagates the cardinality distribution of the multi-object distribution alongside its spatial distribution, eliminating the need to assume a particular form for the cardinality distribution of the estimated densities. The result is that the estimated number of targets is more stable, but it adds to the computational burden of the method and the resulting expressions are more convoluted than the regular PHD filter [57].

More recent efforts have been oriented towards propagating the full multi-object posterior rather than its moments, under the rather general assumption of the process following a multi-Bernoulli distribution [58, 100]. Indeed, multi-Bernoulli distributions do not assume that the multi-object population is IID, and their cardinality distribution is arbitrary.

An important limitation of approaches based on FISST is that as opposed to the classical framework, track identity is not directly propagated in the recursive estimation process due to sets being unordered. An interesting approach to overcoming this limitation was recently proposed alongside a study of conjugate priors for multi-object distributions, where filtering is done on *labeled* RFS [99]. This allows track identity to be preserved through time, but the resulting algorithm is computationally expensive.

2.5.3 Distinguishable stochastic populations

The two previously mentioned multi-object estimation frameworks can be seen as two ends of a spectrum. On one end, the classical approaches propagate tracks and use loose heuristic-based approaches to estimate the configuration of the population; This naturally permits the preservation of track identities but makes population-level modeling harder, the amount of parameters in the track manager larger, and the results less reliable as it is hard to prove filter properties. On the other end, the random finite sets approach estimates the state of the population as a whole, and naturally deals with multi-object estimation issues in a principled way, but loses information on individual track identity due to the underlying set-based assumptions.

In order to solve both these issues, a new multi-object estimation framework was recently proposed where individual track information is preserved, while population dynamics are also modeled. The key idea that is exploited to obtain the best of both worlds is the underlying modeling, which separates *distinguishable* populations, where populations of objects where there is information that uniquely identifies them (e.g., a track that has been observed in the past), and *indistinguishable* populations, where not enough information is available to differentiate objects within them [37]. This is useful to differentiate and estimate together populations such as tracks coming from the population of interest alongside those for tracks which have not yet been observed, and special distributions for clutter, tracks leaving the surveillance region, and so forth.

Estimating *stochastic populations* is in contrast to estimating point processes, as the latter are symmetric and thus do not naturally preserve individual identity. A realization of a stochastic population can be expressed as [14]

$$\mu_{\boldsymbol{n}} = \sum_{i \in \mathbb{I}} \boldsymbol{n}_i \delta[p_i] \tag{2.42}$$

where $\mathbb{I}$ is some indexing set, $\boldsymbol{n}_i$ is the number of objects in population i, δ is the Dirac delta function and p_i is the probability measure indexed by i. If $\boldsymbol{n}_i$ is greater than one, then it is said that objects in population i are indistinguishable. Stochastic populations are random variables that take values such as the one shown above, and these are the mathematical objects which are estimated using this theory.

The Distinguishable Independent Stochastic Population (DISP) filter [14] is based on this paradigm and uses a small set of assumptions in order to propagate as much information as possible about the population of interest. It was designed for challenging scenarios with highly ambiguous data association; for example, when

targets have crossing trajectories or have similar states for long periods of time. In terms of computational complexity, it is similar to the MHT as it propagates a set of *hypotheses*, which are sets of compatible *tracks*. Compatibility here is defined as tracks that can coexist as they have not been associated with the same measurements at any point in their existence. At each time step, every possible association of measurements to either tracks or false alarms is considered, and a new set of hypotheses generated. The DISP filter then computes the probability of each one of these hypotheses. The main advantage with respect to the MHT is that the filtering of tracks and hypotheses is fully probabilistic, and not based in heuristics.

The Hypothesized Independent Stochastic Populations (HISP) filter is based on the same framework but uses an additional assumption in order to greatly reduce its computational requirements. Essentially, where the DISP maintains every possible combination of possible tracks generated by the available observations, the HISP assumes that any two tracks are unlikely to have generated a given observation. This assumption greatly simplifies the data update, making its complexity linear with the number of tracks and the number of measurements.

2.6 Summary

In this chapter, the advantages of performing sequential state estimation in space situational awareness problems were discussed. The challenges that arise when implementing estimation algorithms in this domain were shown, and an overview was given of the approaches that have been used for this purpose in the past.

An overview was given of the most commonly used methods for single- and multi-object estimation. The Bayesian paradigm was described, where it was shown how probability distributions describing the state of a process of interest can be manipulated in order to predict how it evolves through time, and how they can be corrected using data once it becomes available.

In order to tractably estimate these distributions, an appropriate form for the distribution to estimate must be chosen. Each choice produces a different filter, and the most commonly used single-object filtering filters were discussed, including the Kalman filter and its nonlinear variants, solutions based on numerical integration, and sequential Monte Carlo solutions, commonly called particle filters.

A case study where the range of a satellite is measured from a laser ranging facility. It was shown here that non-Gaussian distributions can arise in practical applications, and that careful modeling needs to be done to avoid obtaining inaccurate results.

Several approaches have been pursued to extend this estimation framework to the multi-object case, where it is necessary to estimate the state of multiple objects, and where it is necessary to not only estimate their stochastic state but also the number of present objects. Classical engineering solutions such as the MHT filter were discussed, followed by principled set-based solutions like the PHD filter and finally a new and promising framework which has produced interesting new estimation methods such as the HISP and DISP filters.

Chapter 3

Sensor Modeling for Statistical Orbit Determination

This chapter presents a Bayesian filter designed to estimate the state of Earth-orbiting objects. As it was discussed in chapter 2, propagating the probability distribution of a satellite is essential to having an accurate picture of the uncertainty associated with its known state, and a robust single object filter is necessary to constructing the multi-object estimation algorithms which will be proposed in chapter 4.

In this chapter, different representations for objects in orbit will be analyzed, alongside the models that will be used to predict their position in orbit as time passes. Commonly used sensors for space situational awareness will be modeled, and an analysis of the uncertainty associated with objects that are observed for the first time will be proposed for each sensor.

Having presented the necessary elements to create it, a filter tailored to the estimation of objects in orbit will be presented. This will be based on the concepts presented in chapter 2, and it will be necessary to address the issues that were mentioned there. In particular, a way to generate proposal distributions for sequential Monte Carlo methods will be proposed, after which it will be shown that it will be necessary to devise an approximation for efficient filtering.

The initial orbit determination step of the proposed filter will be based on a particle representation of a uniform distribution, defined on the admissible regions generated by the initial measurements. An update mechanism will be proposed which exploits the linearity of the measurement model in the extended sensor space.

A filter for space situational awareness must be designed with the following operations in mind:

- Determining a prior distribution that represents the initial knowledge of the object state,
- Applying the orbital dynamics model to predict the object's state, and
- Using incoming data from the sensor to reduce the uncertainty of the object state, when available.

Designing the filter will then involve defining a parametrization for the state of the satellites of interest, appropriate dynamical and sensor models, and a way to obtain prior distributions from relevant measurements in the context of recursive Bayesian estimation. The rest of this chapter will begin by addressing these points, followed by their integration into a filter. Finally, results are presented and the chapter concludes.

3.1 State representation and dynamical model

Different parametrizations can be used to represent objects in orbit, and have different uses and advantages at each stage of the filtering process. Cartesian co-ordinates, for instance, are useful to solve the differential equations which model the motion of the Earth-satellite system, and facilitate visualization. Other parametrizations such as sensor-centered spherical co-ordinates can be more useful in order to process sensor measurements. Finally, orbital elements are directly related to orbital dynamics. The Cartesian and orbital element state representations are described below, and the sensor centered spherical co-ordinates will be described in a later section.

3.1.1 Cartesian co-ordinates

When analyzing the motion of objects in orbit of a celestial body, typically a Cartesian co-ordinate system is chosen with origin at the center of the body. When the celestial body of interest is the Earth, geocentric systems are used. Two co-ordinate systems are common: the *eclyptic* system which uses as a fundamental plane the orbital plane of the Earth, and the *equatorial* plane, which uses the plane crossing the equator as fundamental plane. The reference direction x is normally taken to be the *vernal equinox*; that is, when the intersection of the equatorial and eclyptic planes points towards the Sun around the first day of Spring [7]. The reference Cartesian co-ordinate system that will be used in this work is an eclyptic Earth-centered nonrotating inertial (ECI) frame. This space will be denoted $\mathbb{X}^{\mathrm{c}}$. At time step k, the object state $\boldsymbol{x}_k^{\mathrm{c}} \in \mathbb{X}$ is

$$\boldsymbol{x}_k^{\mathrm{c}} = [x_k, y_k, z_k, \dot{x}_k, \dot{y}_k, \dot{z}_k]'. \tag{3.1}$$

The first three elements are the object's position with respect to the reference frame and the last three are its velocity in the same system. Cartesian co-ordinates are simple to visualize and are also used to evaluate important quantities such as the physical distance between two objects, which is necessary to evaluate, for instance, the probability of collision of two objects.

3.1.2 Orbital elements

Orbital elements are a parametrization which arises from the solution of the two-body problem. In the next section, it will be seen how using this parametrization greatly simplifies prediction and modeling the transition kernel, at the cost of an extra transformation. A vector $\boldsymbol{x}_t^{\mathrm{oe}} \in \mathbb{X}^{\mathrm{oe}}$ has the following form:

$$\boldsymbol{x}_t^{\mathrm{oe}} = \begin{bmatrix} \Omega_t \\ \omega_t \\ i_t \\ a_t \\ e_t \\ M_t \end{bmatrix}. \tag{3.2}$$

Here, the *right ascension of the ascending node* Ω_t alongside the *argument of perigee* ω_t and the *inclination* i_t define the orbital plane with respect to the equatorial plane and the reference direction (the vernal equinox). The *semi-major axis* a_t, along with the *eccentricity* e_t, define the shape of the orbit – the first parameter is the scale of the ellipse, whereas the second determines its shape: Circular orbits have zero eccentricity, ellipses have eccentricity between zero and unity, eccentricity greater than one corresponds to hyperbolic orbits, which escape the Earth's gravitational field.. Finally, the *mean anomaly* M_t determines the position of the object along the orbit. More details about orbital elements can be found in [7].

3.1.3 Representation of the distribution

The filtering distributions will be represented using particles. These are essentially samples of the distribution, which can be used to compute expected values from it using Monte Carlo integration [18]. One important difference is that rather than updating these distributions with the typical likelihood weighting process of filters

like the bootstrap filter, the linearity of the measurement model in an alternate parametrization will be exploited to update the distribution with a linear Kalman filter, which will be discussed in Section 3.3.

3.2 Dynamical model

Dynamical models which describe the motion of objects orbiting Earth are usually based on the solution to the differential equation which describes the two body problem for the Earth-satellite system. In its most basic form, this differential equation considers the gravitational force of attraction between the two bodies as the sole acting interaction between them [7]:

$$m_1 \frac{d^2 \boldsymbol{r}_1}{dt^2} = -m_2 \frac{d^2 \boldsymbol{r}_2}{dt^2} = \frac{G m_1 m_2}{r^3} (\boldsymbol{r}_2 - \boldsymbol{r}_1), \tag{3.3}$$

where $m_1, m_2, \boldsymbol{r}_1, \boldsymbol{r}_2$ are the masses and positions of the objects in the system, respectively, G is the gravitational constant, and $r = \|\boldsymbol{r}_1 - \boldsymbol{r}_2\|$. This equation can be simplified if the mass of the orbiting object is considered negligible with comparison to that of Earth, focusing only on the position $\boldsymbol{r}$ of the satellite around Earth:

$$\frac{d^2 \boldsymbol{r}}{dt^2} + \frac{\mu}{r^3} \boldsymbol{r} = 0 \tag{3.4}$$

with $\mu = G(m_E + m_S) \approx G m_E$, where m_E, m_S are the masses of the Earth and the satellite, respectively. Solving this equation analytically involves solving Kepler's equation, which is trascendental. Due to this, solutions to Kepler's problem necessarily involve iterative procedures [7].

Since the two-body problem only considers the gravitational interaction between the satellite and the Earth, it ignores important forces which modify the trajectory of the satellite as it orbits the planet. Typical such sources include third body gravitational forces such as those caused by the sun and the moon, tidal and relativistic effects, changes in atmospheric density, radiation pressure, electromagnetic effects (due to the possible charge of the satellite), self-shadowing, complex unmodeled shakes, numerical integration errors, among others [8]. These sources are usually included in the differential equation through the use of the method of variation of parameters [7].

Very precise numerical integration solutions exist which can be applied to the problem of predicting the future state of an object in orbit. Since the orbital dynamics are very well specified, methods to numerically solve the associated differential

equations, such as the Runge-Kutta family of methods, are widely used to predict the position of satellites over time [7]. Although the proposed particle representation of the state distribution can admit any such method, two methods will be proposed based on the unperturbed two-body problem. First, a simple prediction method based on Shepperd's transition matrix [85] based in Goodyear's solution [28] will be discussed, followed by a dynamical model based on the time evolution of orbital elements. A comparison of the resulting distributions will be presented in the results section.

Shepperd's transition matrix is based only on the unperturbed solution of the two-body problem, and is a function of the previous state of the particle, and the elapsed time between the time steps, Δ_k. Goodyear developed a closed-form transition matrix which, for a given time lapse Δ_t and a given initial state $\boldsymbol{x}_k$, would allow to find the predicted state of the orbiting state, which was subsequently improved by Shepperd by simplifying the required calculations, making its computation simpler. This will be denoted $\Phi(\boldsymbol{x}_k, \Delta_k)$, and equals

$$\Phi(\boldsymbol{x}_k, \Delta_k) = \begin{bmatrix} f & 0 & 0 & g & 0 & 0 \\ 0 & f & 0 & 0 & g & 0 \\ 0 & 0 & f & 0 & 0 & g \\ F & 0 & 0 & G & 0 & 0 \\ 0 & F & 0 & 0 & G & 0 \\ 0 & 0 & F & 0 & 0 & G \end{bmatrix}, \tag{3.5}$$

where f, F, g, and G are obtained from solving Kepler's problem:

$$f = 1 - \frac{\mu_E U_2(w, \beta)}{\|\boldsymbol{r}_0\|} \tag{3.6}$$

$$g = \|\boldsymbol{r}_0\| U_1(w, \beta) + \langle \boldsymbol{r}_0, \boldsymbol{v}_0 \rangle U_2(w, \beta) \tag{3.7}$$

$$F = \frac{-\mu_E U_1(w, \beta)}{(\|\boldsymbol{r}_0\| U_0(w, \beta) + \langle \boldsymbol{r}_0, \boldsymbol{v_0} \rangle U_1(w, \beta) + \mu_E U_2(w, \beta))\|\boldsymbol{r}_0\|} \tag{3.8}$$

$$G = 1 - \frac{\mu_E}{(\|\boldsymbol{r}_0\| U_0(w, \beta) + \langle \boldsymbol{r}_0, \boldsymbol{v}_0 \rangle U_1(w, \beta) + \mu_E U_2(w, \beta)) U_2(w, \beta)}, \tag{3.9}$$

where $\boldsymbol{r}_0, \boldsymbol{v}_0$ are the first and last 3 components of $\boldsymbol{x}_k$, respectively; $\beta = \frac{2\mu_E}{\|\boldsymbol{r}_0\|} - \|\boldsymbol{v}_0\|$ is twice the negative energy of the object; and $U_n(w, \beta)$ are Stumpff's universal functions defined by

$$U_n(w, \beta) = \sum_{k=0}^{\infty} \frac{(-\beta)^k w^{n+2k}}{(n + 2k)!}. \tag{3.10}$$

In order to find the parameter w, the following equation needs to be solved for it:

$$\Delta_k = \|\boldsymbol{r}_0\| U_1(w,\beta) + \langle \boldsymbol{r}_0, \boldsymbol{v}_0 \rangle U_2(w,\beta) + \mu_E U_3(w,\beta). \tag{3.11}$$

Shepperd's approach relies on continued fractions to solve this equation efficiently, and implementation details can be found in his article [85].

The uncertainty in the dynamical model will be modeled as a Gaussian random variable with mean $\mathbf{0}$ and covariance Q_k. This results in the following transition kernel:

$$f(\boldsymbol{x}_k^{\mathrm{c}} | \boldsymbol{x}_{k-1}^{\mathrm{c}}) = \mathcal{N}(\boldsymbol{x}_k^{\mathrm{c}}; \Phi(\boldsymbol{x}_{k-1}^{\mathrm{c}}, \Delta_k)\boldsymbol{x}_{k-1}^{\mathrm{c}}, Q_k), \tag{3.12}$$

where the matrix Q_k represents the uncertainty of the dynamical model due to perturbations and other unmodeled effects. As the dynamics of Earth-orbiting objects are well understood, this term can be very small, and in certain cases an entirely deterministic scenario can be of interest [8]. The impact of this on the filtering process will be discussed in section 3.5.

The dynamical model for orbital elements is simpler:

$$\boldsymbol{x}_t^{\mathrm{oe}} = \boldsymbol{x}_{t-1}^{\mathrm{oe}} + \Delta_t \begin{bmatrix} \mathbf{0}_{5,1} \\ n_{t-1} \end{bmatrix} + \boldsymbol{w}_{t-1} \tag{3.13}$$

where

$$n_{t-1} = \sqrt{\mu_E / a_{t-1}^3} \tag{3.14}$$

is called the *mean motion*, with μ_E the gravitational parameter of the Earth. The vector $\boldsymbol{w}_{t-1} \sim \mathcal{N}(\mathbf{0}_{6,1}, Q_t)$ is a Gaussian random vector which accounts for the uncertainty due to unmodeled effects. The advantage of using this parametrization is that the matrix Q can be chosen such that it reflects the actual uncertainty in the propagation of the elements. For instance, if only uncertainty on the shape of the orbit is desired, noise can be added exclusively on the semi-major axis and eccentricity components. The transformations between orbital elements and Cartesian co-ordinates can be found on [7], and require the solution of Kepler's problem when transforming to Cartesian.

3.3 Sensor modeling

Radars and telescopes are widely used sensors to observe Earth-orbiting objects. Neither sensor provides measurements that can independently determine the full state of the object – Doppler radars usually measure the distance from the sensor

to the measured object, the rate of change of this quantity, and the azimuth and elevation of the object; while optical measurements from telescopes usually comprise azimuth, elevation, and their rates of change by integrating small exposures. Both sensors provide measurements in the topocentric Earth-fixed local horizon frame, so knowledge of the sensor's position and rotation with respect to the Earth-centered reference frame where the object's state is being estimated is required to relate incoming measurements to existing tracks. Computing a measurement produced by an object, then, consists of the following steps:

1. Translate and rotate the object's position so that it is in the sensor's (Cartesian) reference frame,

2. Apply the nonlinear transformation mapping points in the Cartesian reference plane into spherical co-ordinates, and

3. Select the components of the resulting vector which are observed by the sensor.

The rotation and translation are straightforward. Given the time-varying translation vector t_s and rotation matrix W_s from the ECI frame to the sensor's local frame, the position of the satellite in this reference frame, $\tilde{\boldsymbol{x}}_k$, is obtained as

$$\tilde{\boldsymbol{x}}_k = [\tilde{x}_k, \tilde{y}_k, \tilde{z}_k, \dot{\tilde{x}}_k, \dot{\tilde{y}}_k, \dot{\tilde{z}}_k]' = W_\mathrm{s}(\boldsymbol{x}_k - t_\mathrm{s}). \tag{3.15}$$

After this, it is mapped into spherical co-ordinates to obtain a point $\hat{\boldsymbol{z}}_k$ in what will be called the *sensor extended state space* $\mathbb{S}^*$, as it includes both the variables that the sensor observes and those it doesn't. These two operations comprise the transformation $T(\boldsymbol{x}_k, t_\mathrm{s}, W_\mathrm{s})$:

$$\hat{\boldsymbol{z}}_k = T(\boldsymbol{x}_k, t_\mathrm{s}, W_\mathrm{s}) = [r_k, \theta_k, \varphi_k, \dot{r}_k, \dot{\theta}_k, \dot{\varphi}_k]', \tag{3.16}$$

where the individual components are computed the following way:

$$\begin{aligned} r_k &= \sqrt{\tilde{x}_k^2 + \tilde{y}_k^2 + \tilde{z}_k^2} \\ \theta_k &= \text{atan}\frac{\tilde{y}_k}{\tilde{x}_k} \\ \varphi_k &= \text{atan}\frac{\tilde{z}_k}{\sqrt{\tilde{x}_k^2 + \tilde{y}_k^2}} \\ \dot{r}_k &= \frac{\tilde{x}_k\dot{\tilde{x}}_k + \tilde{y}_k\dot{\tilde{y}}_k + \tilde{z}_k\dot{\tilde{z}}_k}{r_k} \\ \dot{\theta}_k &= \frac{\dot{\tilde{y}}_k\tilde{x}_k - \dot{\tilde{x}}_k\tilde{y}_k}{\tilde{x}_k^2 + \tilde{y}_k^2} \\ \dot{\varphi}_k &= \frac{\dot{\tilde{z}}_k(\tilde{x}_k^2 + \tilde{y}_k^2) - \tilde{z}_k(\tilde{x}_k\dot{\tilde{x}}_k + \tilde{y}_k\dot{\tilde{y}}_k)}{r^2\sqrt{\tilde{x}_k^2 + \tilde{y}_k^2}} \end{aligned} \tag{3.17}$$

with $\theta_k \in [-\pi, \pi]$ and $\varphi_k \in [\pi/2, \pi/2]$.

From here, the measurement can be obtained by selecting the components that are measured by the particular sensor which is observing the target:

$$\boldsymbol{z}_k = H\hat{\boldsymbol{z}}_k,$$

where $H = H_\text{r}$ can be used for radars, or $H = H_\text{o}$ for telescopes, with

$$H_\text{r} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad H_\text{o} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In addition to these two sensors, the laser ranging sensor presented in chapter 2 could be used with $H = H_\text{l}$, where

$$H_\text{l} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3.18}$$

Since the focus of this chapter is to provide a full estimation algorithm, and an initial orbit determination method is not available for this type of sensor, however, only the first two types of sensors will be used from now on.

The measurement noise is modeled as additive Gaussian noise, with zero mean and covariance R_k. The covariance will change according to the sensor type and its particular characteristics, but more advanced models can be found in [79] if more

precision is required. The measurement likelihood can then be modeled as

$$g_k(\boldsymbol{z}_k|\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{z}_k - HT(\boldsymbol{x}_k, t_\mathrm{s}, W_\mathrm{s}); \mathbf{0}, R_k) \tag{3.19}$$

It is possible to define models which consider sensor *resolution cells*, and only return measurements in this discrete space. Filtering with this type of measurements is more difficult, however, as the quantization is a nonlinear operation and special filtering techniques need to be used, such as in [19]. In this work, measurements will be assumed to originate in a continuous space.

It is interesting to note here that, if the spherical parametrization $\mathbb{S}^*$ were to be used to represent the object state, a linear Kalman update could be used to update the state and the covariance. This would mean, however, that the time prediction equation would take on a much more difficult form as the topocentric frame is non-inertial, requiring the orbital dynamics to be adjusted accordingly. This remark is the base for the hybrid update mechanism which will be proposed in section 3.5.

3.4 Initial orbit determination

For the types of sensors that are considered in this work, it is possible to constrain the unobserved parameters to specific regions as long as the object is orbiting the Earth [95]. The approach to obtain these constraints and construct priors with them is outlined below for radar and optical sensors.

3.4.1 Radar sensors

For objects with closed orbits around Earth, the internal energy of the object is non-positive:

$$E = \frac{1}{2}\|\dot{\boldsymbol{r}}\|^2 - \frac{\mu}{\|\boldsymbol{r}\|} \leq 0, \tag{3.20}$$

where μ is the mass of the Earth times the gravitational constant, and $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ are the geocentric position and velocity of the satellite. These can be expressed as

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{r}_\mathrm{s} + r\boldsymbol{\rho}_\mathrm{r} \\ \dot{\boldsymbol{r}} &= \dot{\boldsymbol{r}}_\mathrm{s} + \dot{r}\boldsymbol{\rho}_\mathrm{r} + r\dot{\theta}\boldsymbol{\rho}_\theta + r\dot{\varphi}\boldsymbol{\rho}_\varphi, \end{aligned} \tag{3.21}$$

where the position and velocity of the sensor are expressed respectively as $\boldsymbol{r}_\mathrm{s}$ and $\dot{\boldsymbol{r}}_\mathrm{s}$, and

$$\begin{aligned}\boldsymbol{\rho}_\mathrm{r} &= [\cos\theta\cos\varphi, \sin\theta\sin\varphi, \sin\varphi]',\\ \boldsymbol{\rho}_\theta &= [-\sin\theta\cos\varphi, \cos\theta\cos\varphi, 0]', \text{ and}\\ \boldsymbol{\rho}_\varphi &= [-\cos\theta\sin\varphi, -\sin\theta\sin\varphi, \cos\varphi]'.\end{aligned} \tag{3.22}$$

Equation (3.21) is substituted into (3.20) to obtain the following form for the internal energy as a function of the angular rates:

$$\alpha_1\dot{\theta}^2 + \alpha_2\dot{\varphi}^2 + \alpha_3\dot{\theta} + \alpha_4\dot{\varphi} + a_5 \leq 0, \tag{3.23}$$

with

$$\begin{aligned}\alpha_1 &= r^2\cos^2\varphi\\ \alpha_2 &= r^2\\ \alpha_3 &= \frac{r\langle\dot{\boldsymbol{r}}_\mathrm{s}, \boldsymbol{\rho}_\theta\rangle}{2}\\ \alpha_4 &= \frac{r\langle\dot{\boldsymbol{r}}_\mathrm{s}, \boldsymbol{\rho}_\varphi\rangle}{2}\\ \alpha_5 &= \dot{r}^2 + 2\dot{r}\langle\dot{\boldsymbol{r}}_\mathrm{s}, \boldsymbol{\rho}_\mathrm{r}\rangle + \|\dot{\boldsymbol{r}}_\mathrm{s}\|^2 - \frac{2\mu}{Q(r)}, \text{ and}\\ Q(r) &= \sqrt{r^2 + 2r\langle\boldsymbol{r}_\mathrm{s}, \boldsymbol{\rho}_\mathrm{r}\rangle + \|\boldsymbol{r}_\mathrm{s}\|^2}.\end{aligned} \tag{3.24}$$

Following the approach laid out in [95], the admissible values for $(\dot{\theta}, \dot{\phi})$ are those inside of the elliptical disk bounded by

$$\begin{aligned}\dot{\theta} &= \frac{\alpha_3}{\alpha_1} + \sqrt{\frac{\alpha_3^2}{\alpha_1^2} + \frac{\alpha_4^2}{\alpha_1\alpha_2} - \frac{\alpha_5}{\alpha_1}}\cos\phi\\ \dot{\varphi} &= \frac{\alpha_4}{\alpha_2} + \sqrt{\frac{\alpha_4^2}{\alpha_2^2} + \frac{\alpha_3^2}{\alpha_1\alpha_2} - \frac{\alpha_5}{\alpha_2}}\sin\phi,\end{aligned} \tag{3.25}$$

for $\phi \in [0, 2\pi]$. A sample admissible region obtained using this method can be seen in Figure 3.1. Since there is no information to say any point inside this area is more likely than another, the prior distribution is sampled from this area uniformly. To sample from this elliptical disk, a random angle $\phi_s \in [0, 2\pi]$ and range $u_s \in [0, 1]$

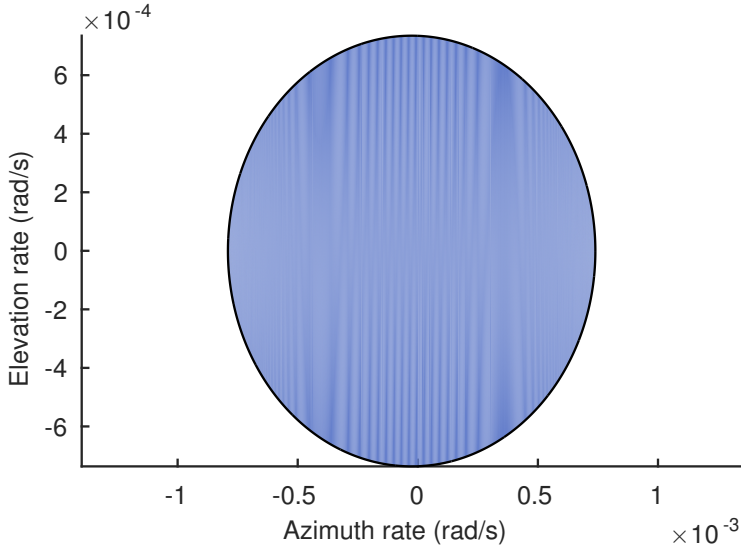


Figure 3.1: Typical angular rates admissible region.

are drawn, after which the sampled $(\dot{\theta}_s, \dot{\phi}_s)$ pair is determined as

$$\begin{aligned} \dot{\theta}_s &= \frac{\alpha_3}{\alpha_1} + \sqrt{u_s}\sqrt{\frac{\alpha_3^2}{\alpha_1^2} + \frac{\alpha_4^2}{\alpha_1\alpha_2} - \frac{\alpha_5}{\alpha_1}}\cos\phi_s, \\ \dot{\varphi}_s &= \frac{\alpha_4}{\alpha_2} + \sqrt{u_s}\sqrt{\frac{\alpha_4^2}{\alpha_2^2} + \frac{\alpha_3^2}{\alpha_1\alpha_2} - \frac{\alpha_5}{\alpha_2}}\sin\phi_s. \end{aligned} \tag{3.26}$$

For each of these samples, a full $\mathbb{S}^*$ vector is constructed by sampling the remaining parameters from the observation likelihood. This process is described in algorithm 3.1.

Algorithm 3.1: Track initialization algorithm using radar measurements

Input :

- Measurement $\boldsymbol{z} = [r, \theta, \varphi, \dot{r}]$
- Covariance for observed parameters R_k

Output: Initialized particle distribution $\{\boldsymbol{x}_0^{(i)}\}_{i=1}^N$
Compute $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ from (3.24)
for $i = 1 \dots N$ **do**
  Sample $\phi_s \sim \mathcal{U}(0, 2\pi)$ and $u_s \sim \mathcal{U}(0, 1)$
  Compute $\dot{\theta}_s, \dot{\varphi}_s$ from (3.26)
  Sample $[r_s, \theta_s, \varphi_s, \dot{r}_s]' \sim \mathcal{N}(\cdot, z, R_k)$
  Set $\boldsymbol{x}_0^{(i)} \leftarrow T^{-1}([r_s, \theta_s, \varphi_s, \dot{r}_s, \dot{\theta}_s, \dot{\varphi}_s]')$
end

3.4.2 Optical sensors

Although many different types of optical sensors exist, the focus here will be in those that can measure angles of observation (e.g., azimuth and elevation) and their rates of change. As before, the two unobserved parameters can be bounded by using energy constraints on the object being tracked. Equation (3.21) is substituted into (3.20) to obtain the energy in terms of range and range rate:

$$2E = \dot{r}^2 + w_1\dot{r} + T(r) - \frac{2\mu}{\sqrt{S(r)}} < 0, \tag{3.27}$$

where

$$\begin{aligned} S(r) &= r^2 + w_5 r + w_0, \\ T(r) &= w_2 r^2 + w_3 r + w_4, \\ w_0 &= \|\boldsymbol{r}_\mathrm{s}\|^2, \\ w_1 &= 2\langle \dot{\boldsymbol{r}}_\mathrm{s}, \boldsymbol{\rho}_\mathrm{r} \rangle, \\ w_2 &= \dot{\theta}^2 \cos^2\varphi + \dot{\varphi}^2, \\ w_3 &= 2\langle \boldsymbol{r}_\mathrm{s}, \dot{\theta}\boldsymbol{\rho}_\theta + \dot{\varphi}\boldsymbol{\rho}_\varphi \rangle, \\ w_4 &= \|\dot{\boldsymbol{r}}\|^2, \text{ and} \\ w_5 &= 2\langle \boldsymbol{r}_\mathrm{s}, \boldsymbol{\rho}_\mathrm{r} \rangle. \end{aligned}$$

From (3.27), if a value for r is given, then $\dot{r}$ is constrained to the following interval:

$$\dot{r} \in (-w_1/2 - \zeta(r), -w_1/2 + \zeta(r)) \tag{3.28}$$

$$\zeta(r) = \frac{1}{2}\sqrt{w_1^2 - 4(T(r) - 2\mu/\sqrt{S(r)})} \tag{3.29}$$

A typical admissible region obtained this way can be seen in Figure 3.2. As an additional constraint, minimum and maximum ranges can be defined for objects of interest: $r \in [r_\mathrm{min}, r_\mathrm{max}]$. To sample uniformly from this region, rejection sampling is used: The maximum and minimum values for r and $\dot{r}$ are determined, and points are sampled uniformly from this rectangle. Equation (3.28) is used to verify this is a valid solution; If it is, it is admitted into the group of samples. If it is not, the process is repeated until the amount of required samples is obtained. The algorithm is detailed in listing 3.2.

The obtained admissible region can be made smaller if additional constraints are posed on the eccentricity and the semimajor axis following the method proposed

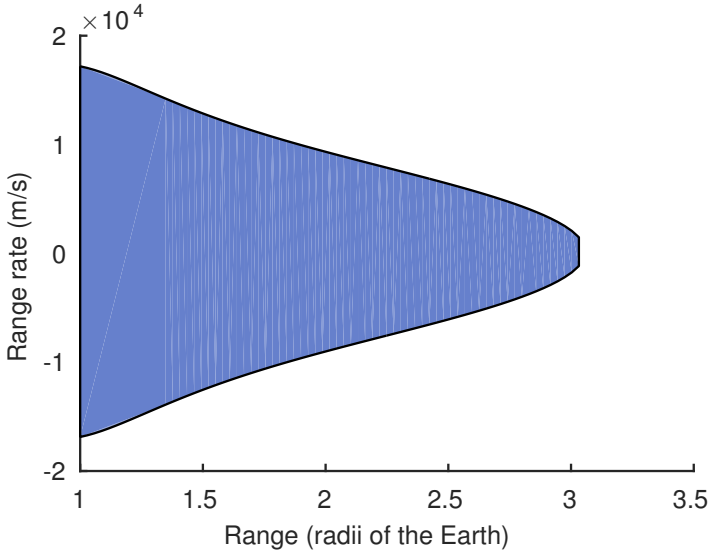


Figure 3.2: Typical range – range rate admissible region.

in [16], e.g., if the objects of interest are in specific orbits. However, since no prior knowledge is assumed on the orbits of the observed objects these additional constraints will not be used here.

3.5 Filtering recursion

A suitable filter to estimate the state of Earth-orbiting satellites must be able to use priors of the form described in section 3.4, in addition to the nonlinear dynamical and observation models outlined above. The nonlinear, non Gaussian nature of this problem makes Kalman filters and their nonlinear variants unfit for this purpose. The focus will be on designing a particle filter which can accurately propagate the state distributions through time, and correct them as measurements are acquired.

Although SIR filters are very flexible, a problem arises when the Markov transition kernel (3.12) has low covariance. This is due to the fact that when using a particle representation in (2.2), the weight update becomes

$$w_k^{(i)} = \frac{\tilde{w}_k^{(i)}}{\sum_{j=1}^{N} \tilde{w}_j^{(i)}}, \tag{3.30}$$

where

$$\tilde{w}_k^{(i)} = w_{k-1}^{(i)} \frac{f(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)}) g_k(\boldsymbol{z}_k|\boldsymbol{x}_k^{(i)})}{\pi(\boldsymbol{x}_k^{(i)})}, \tag{3.31}$$

Algorithm 3.2: Track initialization algorithm using optical measurements

Data:

- Measurement $\boldsymbol{z} = [\theta, \varphi, \dot{\theta}, \dot{\varphi}]$
- Range limits $r_{\min}, r_{\max}$
- Covariance for observed parameters R_k

Result: Initialized particle distribution $\{\boldsymbol{x}_0^{(i)}\}_{i=1}^N$
Evaluate $w_0, w_1, w_2, w_3, w_4, w_5, T(r), S(r)$
Compute $\dot{r}_{\min} = -w_1/2 - \zeta(r_{\min})$ and $\dot{r}_{\max} = -w_1/2 + \zeta(r_{\min})$
for $i = 1 \dots N$ **do**
    solution_found $\leftarrow$ `false`
    while not solution_found **do**
        Sample $r^{(i)} \sim \mathcal{U}(r_{\min}, r_{\max})$
        Sample $\dot{r}^{(i)} \sim \mathcal{U}(\dot{r}_{\min}, \dot{r}_{\max})$
        Evaluate ζ from (3.29)
        if ζ *is real and* (3.28) *holds* **then**
            solution_found $\leftarrow$ `true`
        end
    end
    Sample $[\theta^{(i)}, \varphi^{(i)}, \dot{\theta}^{(i)}, \dot{\varphi}^{(i)}]' \sim \mathcal{N}(\cdot; z, R)$
    $\boldsymbol{x}_0^{(i)} \leftarrow T^{-1}([r^{(i)}, \theta^{(i)}, \varphi^{(i)}, \dot{r}^{(i)}, \dot{\theta}^{(i)}, \dot{\varphi}^{(i)}]')$
end

which will make the weights quickly go to zero if the newly sampled values $\boldsymbol{x}_k^{(i)}$ are unlikely under the transition kernel $f(\cdot|\boldsymbol{x}_{k-1}^{(i)})$. At the beginning of the filtering process, where there is high uncertainty on the target position and velocities induced by the shape of the distribution yielded by the initial orbit determination, this is bound to happen often. There are two particular forms of the importance sampling function π which avoid this problem. The bootstrap filter uses $\pi(\boldsymbol{x}_k^{(i)}) = f(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})$, which cancels out in (3.31) with the problematic part in the numerator. This proposal, however, does not use measurement information and is bound to be less efficient than a fully adapted proposal distribution [18]. Another option is to use the optimal proposal function $\pi(\boldsymbol{x}_k^{(i)}) = f(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})g_k(\boldsymbol{z}_k|\boldsymbol{x}_k^{(i)})$, in which case the weights would remain constant through the filtering duration. In a typical SSA application, this distribution is not available for direct sampling, due to the form of the measurement and dynamical models. An approximation of this distribution based on the linearity of the measurement model in the sensor state space will be outlined below.

Although the dynamical model discussed in section 3.2 is a linear, Gaussian approximation in the object's state space (conditioned a given previous state) and the sensor model is linear and Gaussian in the extended sensor state space discussed in section 3.3, the nonlinear transformation between the two spaces means that a closed form for the updated distribution is not available.

Using a particle representation for the object state provides a straightforward way to map between the two spaces, by simply applying the required transformation to each particle. This is an essential part of the proposed method, whether for radar or optical measurements. The particle distribution is updated as measurements arrive by approximating it as a Gaussian distribution in $\mathbb{S}^*$, and then applying a Kalman update to it before sampling a new set of particles from the updated distribution. This approach is outlined below. To begin, the unweighted particles from the previous time step $\{\boldsymbol{x}_{k-1}^{(i)}\}_{i=1}^N$ are used to produce a set of predicted particles by drawing from the transition kernel (3.12):

$$\boldsymbol{x}_{k|k-1}^{(i)} \sim \mathcal{N}(\cdot; \Phi(\boldsymbol{x}_{k-1}, \Delta_k)\boldsymbol{x}_{k-1}, Q_k).$$

If no measurement is available, then these particles are kept as the predicted distribution. However, if a measurement has been received, the particles are first mapped to $\mathbb{S}^*$ using (3.17) to obtain

$$\boldsymbol{y}_{k|k-1}^{(i)} = T(\boldsymbol{x}_{k|k-1}^{(i)}, t_\mathrm{s}, W_\mathrm{s}).$$

The mean and covariance of this set of particles are computed:

$$\boldsymbol{\mu}_{k|k-1} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_{k|k-1}^{(i)},$$
$$\Sigma_{k|k-1} = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{y}_{k|k-1}^{(i)} - \boldsymbol{\mu}_{k|k-1})^2,$$

after which a linear Kalman Filter update is performed [47]:

$$\begin{aligned} S &= H\Sigma_{k|k-1}H' + R, \\ K &= \Sigma_{k|k-1}H'S^{-1}, \\ \boldsymbol{\mu}_k &= \boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z}_k - H\boldsymbol{\mu}_{k|k-1}), \\ \Sigma_k &= (I - KH)\Sigma_{k|k-1}. \end{aligned} \tag{3.32}$$

Here, S denotes the innovation covariance and K the Kalman gain. The updated set of particles is drawn from a Gaussian distribution with these parameters

$$\boldsymbol{y}_k^{(i)} \sim \mathcal{N}(\cdot; \boldsymbol{\mu}_k, \Sigma_k), \tag{3.33}$$

and finally the updated set of particles is obtained by mapping back to $\mathbb{X}$,

$$\boldsymbol{x}_k^{(i)} = T^{-1}(\boldsymbol{y}_k^{(i)}, t_\mathrm{s}, W_\mathrm{s}). \tag{3.34}$$

Although this approach has been shown to perform very well, it is important to note that it is not strictly a particle filtering method as the distribution is not weighted. Two concerns with this approach are that approximating the distribution as a Gaussian may cause some loss of information on the shape of the distribution, and sampling from a single distribution discards the particle trajectories. Below, an importance sampling function is sketched which reproduces this hybrid update approach while maintaining the theoretical properties of a SIR filter. To do this, it can be remarked that the correction step can be applied to each individual particle, rather than the distribution mean, to obtain

$$\boldsymbol{y}_k^{(i)} = \boldsymbol{y}_{k|k-1}^{(i)} + K(\boldsymbol{z}_k - H\boldsymbol{y}_{k|k-1}^{(i)}) \tag{3.35}$$

the mean of this distribution is the following:

$$\begin{aligned} \mathrm{E}[\boldsymbol{y}_k^{(i)}] &= \mathrm{E}[\boldsymbol{y}_{k|k-1}^{(i)}] + K(\boldsymbol{z}_k - H\mathrm{E}[\boldsymbol{y}_{k|k-1}^{(i)}]) \\ &= \boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z}_k - H\boldsymbol{\mu}_{k|k-1}), \end{aligned} \tag{3.36}$$

which matches the mean of (3.33). The covariance is given by

$$\begin{aligned} \mathrm{cov}[\boldsymbol{y}_k^{(i)}] &= (I - KH)\Sigma_{k|k-1}(I - KH)' \\ &= (I - KH)\Sigma_{k|k-1} - (\Sigma_{k|k-1}H'K' - KH\Sigma_{k|k-1}H'K') \\ &= (I - KH)\Sigma_{k|k-1} - (\Sigma_{k|k-1}H'K' - K(S - R)K') \\ &= (I - KH)\Sigma_{k|k-1} - (\Sigma_{k|k-1}H'K' - KSK' + KRK') \\ &= (I - KH)\Sigma_{k|k-1} - (\Sigma_{k|k-1}H'K' - \Sigma_{k|k-1}H'K' + KRK') \\ &= (I - KH)\Sigma_{k|k-1} - KRK'. \end{aligned} \tag{3.37}$$

This means that the covariance will be underestimated by comparison to the previous method by a factor of KRK'. In order to compensate for this, an independent random variable with covariance KRK' can be added such that the total variance will be $(I - KH)\Sigma_{k|k-1}$:

$$\boldsymbol{y}_k^{(i)} = \boldsymbol{y}_{k|k-1}^{(i)} + K(\boldsymbol{z}_k - H\boldsymbol{y}_{k|k-1}^{(i)}) + \boldsymbol{\nu}_k \tag{3.38}$$

where $\boldsymbol{\nu}_k^{(i)} \sim \mathcal{N}(\cdot; 0, KRK')$, for example. In this case, the value of the importance sampling function would be

$$\pi(\boldsymbol{y}_k^{(i)}) = \mathcal{N}(\boldsymbol{\nu}_k^{(i)}; 0, KRK'))\mathcal{N}(\boldsymbol{x}_{k|k-1}^{(i)}; \Phi(\boldsymbol{x}_{k-1}, \Delta_k)\boldsymbol{x}_{k-1}, Q_k). \tag{3.39}$$

An alternative is to inject variance into the resulting distribution by uniformly moving the particles away from the mean. In this case, for a given matrix F, the particles would be computed as

$$\boldsymbol{y}_k^{(i)} = (I + F)(\boldsymbol{y}_{k|k-1}^{(i)} + K(\boldsymbol{z}_k - H\boldsymbol{y}_{k|k-1}^{(i)})) - F\boldsymbol{\mu}_k, \tag{3.40}$$

in which case the mean would be

$$\begin{aligned} \mathrm{E}[\boldsymbol{y}_k^{(i)}] &= (I + F)(\mathrm{E}[\boldsymbol{y}_{k|k-1}^{(i)}] + K(\boldsymbol{z}_k - H\mathrm{E}[\boldsymbol{y}_{k|k-1}^{(i)}])) - F\mu_k \\ &= (I + F)(\boldsymbol{\mu}_{k|k-1} + K(\boldsymbol{z}_k - H\boldsymbol{\mu}_{k|k-1})) - F\boldsymbol{\mu}_k \\ &= \boldsymbol{\mu}_k + F\boldsymbol{\mu}_k - F\boldsymbol{\mu}_k \\ &= \boldsymbol{\mu}_k, \end{aligned} \tag{3.41}$$

which is as required, and the covariance would be

$$\begin{aligned}\operatorname{cov}[\boldsymbol{y}_k^{(i)}] &= (I+F)(I-KH)\Sigma_{k|k-1}(I-KH)'(I+F)' \\ &= (I+F)(\Sigma^{\vee})(I+F)' \\ &= \Sigma^{\vee} + F\Sigma^{\vee} + \Sigma^{\vee}F' + F\Sigma^{\vee}F',\end{aligned} \tag{3.42}$$

where $\Sigma^{\vee} = (I-KH)\Sigma_{k|k-1} - KRK'$. To get the required covariance, F must be a solution to

$$\Sigma^{\vee} + F\Sigma^{\vee} + \Sigma^{\vee}F' + F\Sigma^{\vee}F' = KRK'. \tag{3.43}$$

If a symmetry condition is imposed on F, this equation can be solved using the method described in [67]. In this case, the importance function can be evaluated as

$$\pi(y_k^{(i)}) = \mathcal{N}(x_{k|k-1}^{(i)}; \Phi(\boldsymbol{x}_{k-1}, \Delta_k)\boldsymbol{x}_{k-1}, Q_k). \tag{3.44}$$

Unfortunately, these importance sampling approaches have the problem that the numerator of (3.31) will be either zero or very close to it as the kernel $f(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})$ will usually be very narrow (or in the deterministic case just a Dirac delta) compared to the correction applied by the Kalman update, causing degenerate distributions as most weights will go to 0. For this reason, the particle set obtained from (3.33) will be used instead. This can be seen as an approximation of the optimal importance sampling function as the sample is unweighted. A flowchart detailing the filtering process can be seen in Figure 3.3. The prediction and update algorithms can be seen in Algorithms 3.3 and 3.4, respectively.

Algorithm 3.3: Prediction algorithm

Data:

- Particle distribution $\{\boldsymbol{x}_{k-1}^{(i)}\}_{i=1}^{N}$ at time step $k-1$
- Covariance matrix of the process noise Q_k
- Elapsed amount of time Δ_k
- Translation and rotation to sensor Cartesian frame $t_{\mathrm{s}}, W_{\mathrm{s}}$

Result: Predicted particle distribution $\{\boldsymbol{x}_{k|k-1}^{(i)}\}_{i=1}^{N}$

for $i = 1 \dots N$ **do**

Sample $\epsilon_i \sim N(\mathbf{0}, Q_k)$

$\Phi \leftarrow$ `Shepperd_matrix`$(\boldsymbol{x}_{k-1}^{(i)}, \Delta_k)$

$\boldsymbol{x}_{k|k-1}^{(i)} \leftarrow \Phi\boldsymbol{x}_{k-1}^{(i)} + \epsilon_i$

end

Algorithm 3.4: Update algorithm

Data:

- Predicted particle distribution $\{\boldsymbol{x}_{k|k-1}^{(i)}\}_{i=1}^{N}$ at time step k
- Radar or optical measurement $\boldsymbol{z}$
- Covariance matrix of the noise R_k

Result: Updated particle distribution $\{\boldsymbol{x}_{k}^{(i)}\}_{i=1}^{N}$

Choose H according to the type of sensor to use

for $i = 1 \dots N$ **do**

$\quad \boldsymbol{y}_i^- \leftarrow T(\boldsymbol{x}_{k|k-1}^{(i)}, t_\mathrm{s}, R_\mathrm{s})$

end

$\boldsymbol{\mu}_k^- \leftarrow \mathrm{mean}(\{\boldsymbol{y}_i^-\}_{i=1}^{N})$

$\Sigma_k^- \leftarrow \mathrm{cov}(\{\boldsymbol{y}_i^-\}_{i=1}^{N})$

Compute Kalman Filter updated mean and covariance

$\boldsymbol{\xi} \leftarrow \boldsymbol{z} - H\boldsymbol{\mu}_k^-$

$S \leftarrow H\Sigma_k^- H' + R_k$

$K \leftarrow \Sigma_k^- H' S^{-1}$

$\boldsymbol{\mu}_k \leftarrow \boldsymbol{\mu}_k^- + K\boldsymbol{\xi}$

$\Sigma_k \leftarrow (I - KH)\Sigma_k^-$

for $i = 1 \dots N$ **do**

$\quad$ Sample $\boldsymbol{y}_i \sim N(\boldsymbol{\mu}_k, \Sigma_k)$

$\quad \boldsymbol{x}_k^{(i)} \leftarrow T^{-1}(\boldsymbol{y}_i, t_\mathrm{s}, R_\mathrm{s}))$

end

3.5.1 Performing inference

Once the filtering distributions have been approximated with the algorithms described above, they can be used to compute useful information such as the expected position of a satellite, or to approximate the probability of collision between two bodies. In general, performing inference with Monte Carlo methods is very flexible; it could be interesting to know, for instance, the confidence intervals of the obtained estimators or the highest posterior density regions of the obtained distribution [42]. However, these advanced inference techniques are out of the scope of this thesis and are left for future work.

Obtaining a point estimate from the obtained distributions can be useful to indicate the likeliest state given the available data. As it was previously discussed, the distributions are made up of unweighted particles which means that the Maximum A Posteriori (MAP) estimate used commonly in particle filtering approaches, which

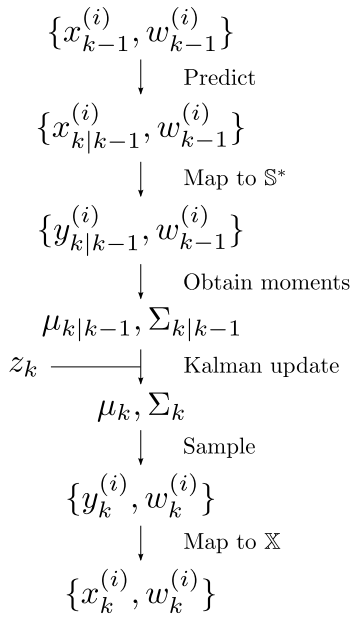


Figure 3.3: Flowchart for the filtering algorithm with global correction

returns the state of the particle with the highest weight, is not suitable in this case.

Due to the nature of orbital mechanics, the spread of particles after prediction will tend to lay along the same orbit and thus take a banana-shaped distribution. Taking the Expected A Posteriori (EAP) estimate from the distribution in the Cartesian space $\mathbb{X}$, then, will introduce bias in the estimated orbit as illustrated in Figure 3.4 for a sample cloud of particles. To solve this, it is proposed to first map the distribution into the augmented sensor frame $\mathbb{S}^*$, compute the EAP in this space, and then map this value back into $\mathbb{X}$ as the filter output. The resulting estimator is compared to the original one in Figure 3.4, where it can be seen that the filter estimate agrees more with the estimated orbit as predicted by the orbital dynamics. The method is described in Algorithm 3.5.

Algorithm 3.5: State estimate extraction process

Data:

- Particle distribution $\{\boldsymbol{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ at time step k

Result: Estimated state $\hat{\boldsymbol{x}}_k$
for $i = 1 \dots N$ **do**
    $\hat{\boldsymbol{y}}_i \leftarrow T(\boldsymbol{x}_k^{(i)}, t_\mathrm{s}, R_\mathrm{s})$
end
$\mu_k \leftarrow \texttt{mean}(\{\boldsymbol{y}_i\}_{i=1}^N)$
$\hat{x}_k \leftarrow T^{-1}(\boldsymbol{\mu}_k, t_\mathrm{s}, R_\mathrm{s}))$

To approximate the probability of collision, the expected value of a function

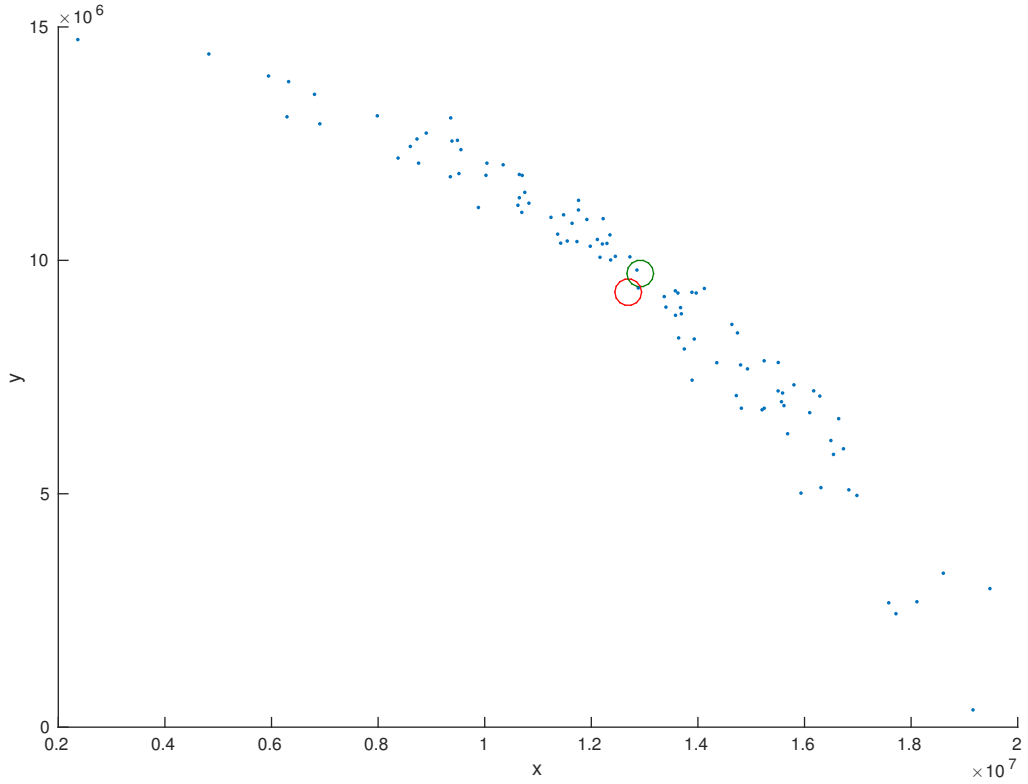


Figure 3.4: EAP of the Cartesian distribution (red circle) compared to the proposed state estimate (green circle).

determining if the two objects are closer than a distance representing their size could be computed. This assumes that the objects are spherical and have the same size, but the distance parameter could be adjusted to relax this assumption. If the function

$$\text{collision}(x, y) = \begin{cases} 1, & \text{if } d(x, y) < T, \\ 0 & \text{otherwise,} \end{cases} \tag{3.45}$$

with $d(x, y)$ the distance between points x and y in Cartesian co-ordinates, then the probability of collision between objects $X_{1,k}, X_{2,k}$ with particle distributions $\{\boldsymbol{x}_{1,k}^{(i)}, w_{1,k}^{(i)}\}_{i=1}^{N}, \{\boldsymbol{x}_{2,k}^{(i)}, w_{2,k}^{(i)}\}_{i=1}^{N}$, respectively, could be approximated as its expected value:

$$\begin{aligned} p_{\text{c}} &= \text{E}[\text{collision}(X_1, X_2)] \\ &\approx \sum_{i=1}^{N} \sum_{j=1}^{N} w_{1,k}^{(i)} w_{2,k}^{(j)} \, \text{collision}(\boldsymbol{x}_{1,k}^{(i)}, \boldsymbol{x}_{2,k}^{(j)}). \end{aligned}$$

3.6 Experiments

In this section, an analysis is made of the validity of the assumptions of the proposed algorithm, followed by experiments on simulated data, and a brief discussion about how this method compares to other filtering algorithms is finally presented.

3.6.1 Comparison to the true posterior

This chapter presents an approximation of the filtering distribution obtained from Bayes' filter,

$$p(\boldsymbol{x}_k) = \frac{f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})g_k(\boldsymbol{z}_k|\boldsymbol{x}_k)}{\int f(\boldsymbol{x}'_k|\boldsymbol{x}_{k-1})g_k(\boldsymbol{z}_k|\boldsymbol{x}'_k)\,d\boldsymbol{x}'_k}. \tag{3.46}$$

Although as it was discussed before there is no closed form for this distribution, simulation methods such as the ones discussed in chapter 2 could be used to sample from the true posterior density. This is beneficial as it can be used to analyze how faithfully the approximated distribution resembles the real posterior.

In order to obtain samples from the posterior, the Metropolis-Hastings algorithm was used using the denominator of (3.46) and a Gaussian kernel. Due to the problems described in section 3.5, however, sampling directly from this distribution was problematic. The main problem was that it was hard for the term corresponding to the Markov transition kernel to be greater than zero, as it is very narrow compared to both the measurement likelihood and the prior, so all of the samples were rejected and the algorithm would be stuck at the initial value.

Although more sophisticated sampling algorithms exist, where for instance bridging densities are used to help the MCMC method converge smoothly to the target distribution, these usually involve a higher computational burden [18]. Since it would still be interesting to compare the approximation done here with the true posterior, however, the implementation of these will be left for future work.

3.6.2 Analysis of the distributions

The Kalman Filter provides an estimator which is optimal in the sense of minimal variance if the prior distribution is Gaussian and the measurement model is linear. The distributions which are derived from the initial orbit determination methods described above are not, however, Gaussian, and thus using a Kalman filter naively can provide invalid results.

Gaussian distributions are completely characterized by their first two moments, but the uniform distributions on the admissible regions described before need more

parameters to be characterized. The Gaussian approximation that is done, then, truncates some of the available information and can generate samples that are not admissible. However, approximating this distribution as a Gaussian yields big performance gains since the number of required particles is relatively low, as discussed in section 3.5. Fortunately, the generated distributions are not multimodal as they are composed of Gaussian and uniform components, and so the Gaussian approximation is reasonable. The uniform components could be better approximated with distributions with fatter tails, such as Student's t distribution, and so filters for these such as the one in [77] could reasonably be used here. Alternatively, Gaussian mixtures such as the ones described in chapter 2 could be used to increase the faithfulness of the approximation, as in [16]. The Kalman filter, however, is both simple and efficient compared to these two approaches so it will be used in this work.

Since the filtering distribution is approximated as a Gaussian when there is an update step, it is important to see how robust the filter is to these reparametrizations as they can entail some information loss. To do this, a Gaussian distribution in the sensor state space will be generated and non-linear transformations of two types will be applied to it:

- First, when the relative position of the object in the sensor frame of reference $\mathbb{S}^*$ evolves;
- Second, when the spatial distribution of the object in the Earth frame of reference $\mathbb{X}$ is corrupted with noise.

In order to evaluate if a distribution is still Gaussian after each of these transformations, Henze and Zirkler's BHEP test [35] is used. This test compares the theoretical characteristic function of a Gaussian distribution with the empirical characteristic function of the available samples, and develops a test statistic which can be used to test whether it is plausible for the samples to have come from a Gaussian distribution. The p-values of the test that are presented below indicate the probability of the test statistic (or a more extreme result) having been produce under the hypothesis of the distribution being Gaussian. The hypothesis is rejected if the p-value falls under a specified threshold probability, e.g., 0.05.

In the first experiment a Gaussian distribution in spherical co-ordinates is initialized, corresponding to an object directly above a sensor and at a distance of 4 times the radius of the Earth, with covariance

$$\text{diag}([1000^2, (\pi/180)^2, (\pi/180)^2, 100^2, (\pi/180)^2, (\pi/180)^2]),$$

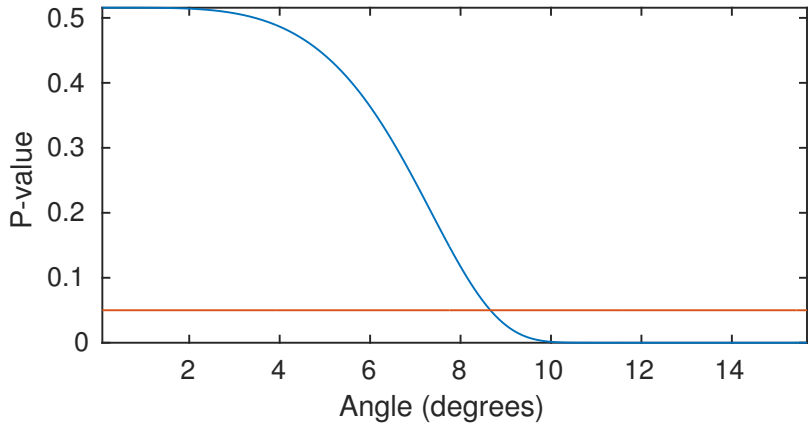


Figure 3.5: Results of the BHEP test averaged over 100 Monte Carlo runs for the first sensibility experiment. The red line indicates the 0.05 confidence interval, under which the hypothesis of the distribution being Gaussian is rejected.

where diag(·) denotes the diagonal matrix with elements in the diagonal given by its argument. The sensor frame of reference is then modified, corresponding to the sensor motion induced by the rotation of the Earth during 20 seconds, and the initial distribution is mapped to this new co-ordinate frame. The BHEP test is used to analyze the probability of the distribution being Gaussian in this new frame, and the process is iterated for a duration of approximately 1 hour. This was repeated for 100 Monte Carlo runs, and the average values of the p-values yielded by the test can be seen in Figure 3.5 graphed against the relative angle to the initial sensor frame of reference.

The results of the first experiment suggest that the validity of the Gaussian assumption in a sensor frame of reference is robust to moderate alterations of the sensor frame of reference. While out of the scope of this chapter, it opens the possibility for the exploitation of the sensor co-ordinate parameterization when two sensors observe simultaneously the same orbiting object.

In the second experiment a Gaussian distribution in spherical co-ordinates is initialized as in the first one. The spatial distribution is then mapped to the Cartesian frame of reference, corrupted with Gaussian noise in this parametrization with covariance

$$\text{diag}([1000^2, 1000^2, 1000^2, 100^2, 100^2, 100^2]),$$

before being mapped back to sensor frame of reference. The BHEP test is then used to assert whether the resulting distribution is Gaussian. This procedure is then repeated for growing levels of noise. The experiment was performed on 100 Monte Carlo runs, and the averaged p-values can be seen in Figure 3.6.

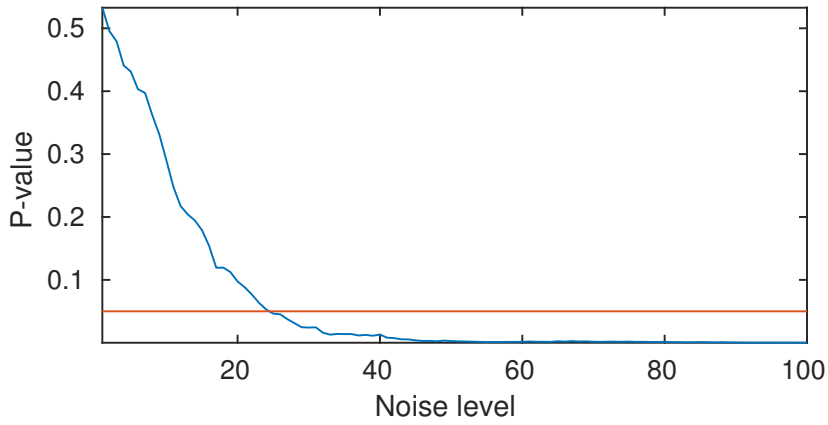


Figure 3.6: Results of the BHEP test averaged over 100 Monte Carlo runs for the second sensibility experiment. The noise level indicates how many times Gaussian noise was added to the distribution in Cartesian co-ordinates. The red line indicates the 0.05 confidence interval.

T.	a	e	i	Ω	ω
1	16495.0 km	0.01	2.0°	20.0°	311.0°
2	26352.5 km	0.6	11.3°	60.0°	351.0°

Table 3.1: Target index, semi-major axis (a), eccentricity (e), inclination (i), right ascension of the ascending node (Ω), and argument of perigee (ω).

The results of the first experiment suggest that the validity of the Gaussian assumption in a sensor frame of reference is robust to the corruption of the distribution in the Cartesian frame of reference with significant noise levels. In particular, since the level of the process noise Q_k in the prediction step (3.12) is significantly lower than the threshold over which the BHEP test fails in Figure 3.6, these results suggest that the validity of the Gaussian approximation hold for successive filtering steps while an object is in the sensor's field of view and thus frequently observed and updated.

3.6.3 Filtering results

The performance of the proposed filter will be evaluated on two simulated orbits which were generated using the method described in [26], where a Runge-Kutta 7/8 numerical integration method is used to propagate the orbits of the objects taking into account the gravitational field of the Earth up to order and degree 12, including third-body perturbations of the Sun and the Moon and radiation pressure. The simulated objects are spherical and have area-to-mass ratios of $0.02\text{m}^2/\text{kg}$ and their orbital characteristics can be seen in Table 3.1.

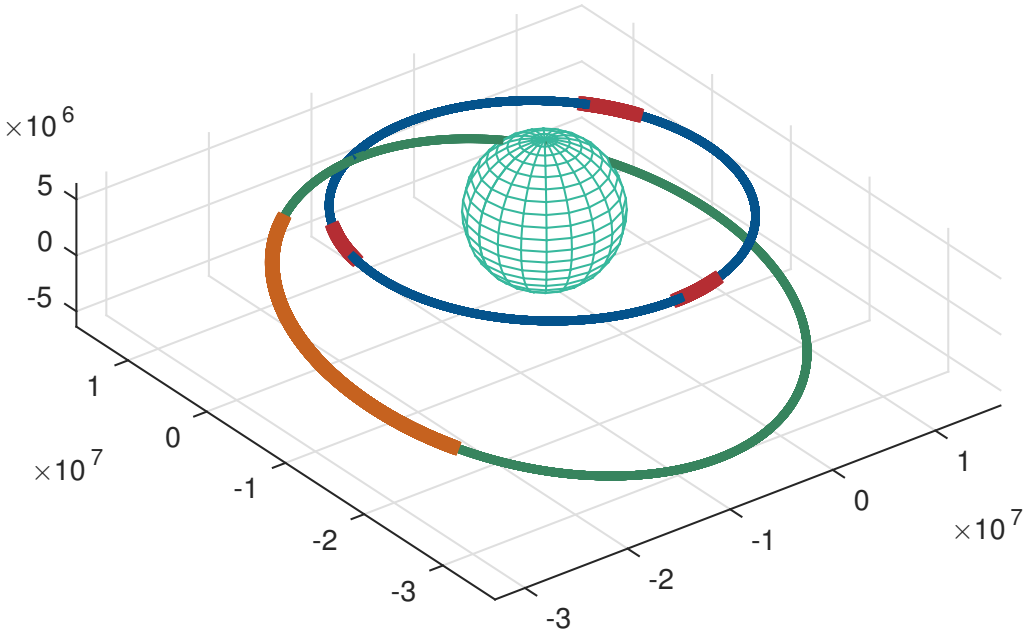


Figure 3.7: Orbits used for performance evaluation. Target 1 is shown in red when observed and blue otherwise; target 2 in orange when observed and green otherwise.

An illustration of the orbits can be seen in Figure 3.7, including the periods where they are observed by the Earth-based sensors. The challenges for each filtering scenario are different - Target 1 is only seen for brief periods of time as it flies past the sensor field of view, while Target 2 is seen for an uninterrupted time interval. For each target, 100 sets of measurements are simulated for either a radar (Target 1) or a telescope (Target 2), and the estimation error is calculated for each.

Comparison of dynamical models

The dynamical model will be chosen to be the one based on orbital element propagation, as it is possible to have a model of dynamical noise which reflects more closely what is expected in orbital dynamics. To illustrate this, the described filter was run on the radar measurements produced by target one. The measured covariance in Earth-centred spherical co-ordinates can be seen in figures 3.8 and 3.9, shown for each parameter. It is clear here that the orbital elements parametrization can be used to add a reasonable amount of covariance during filtering, without adding too much uncertainty in the periods when the target is not observed.

Simulations

The first target is a low-earth orbit satellite which is observed for three brief time intervals as it comes into the field of view of the sensor. A radar is used for this

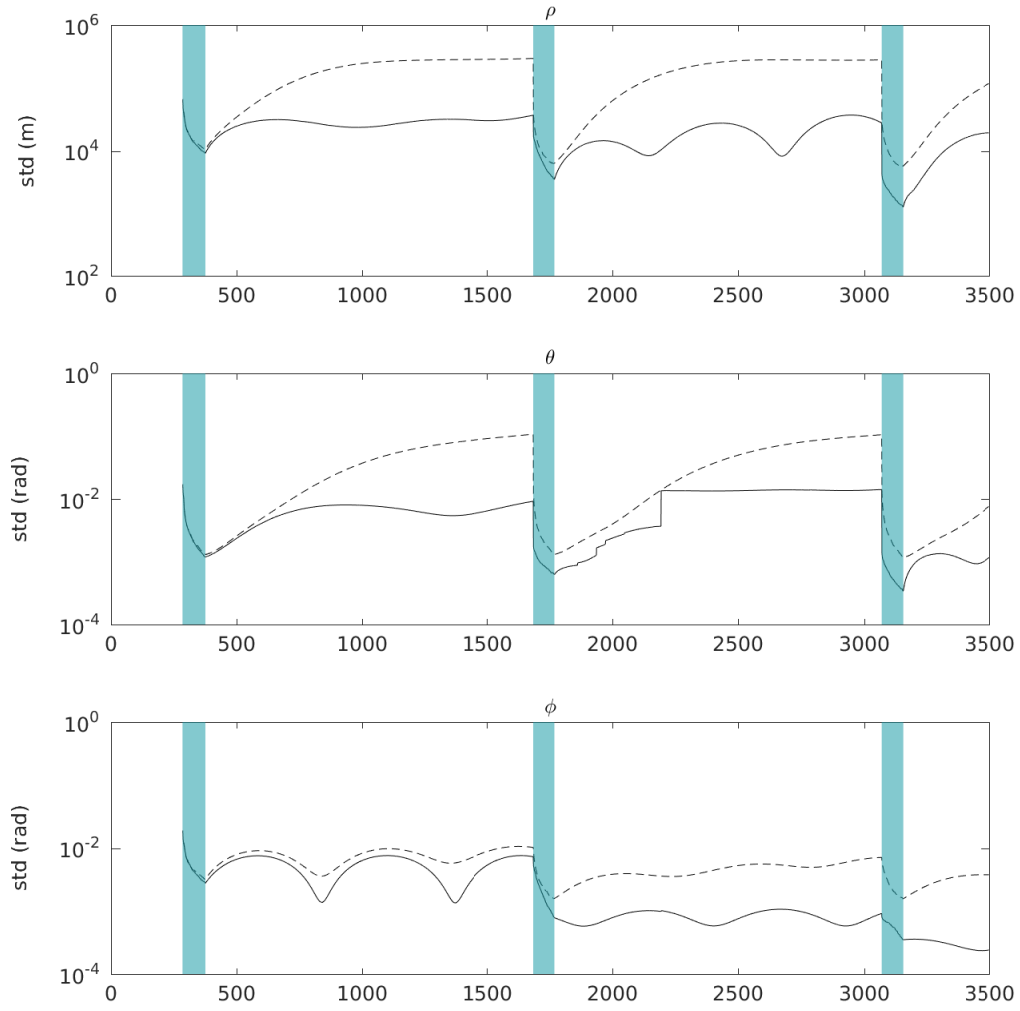


Figure 3.8: Estimated standard deviation for range, azimuth, and elevation, for the orbital elements prediction (continuous line) and the Shepperd matrix prediction (dashed line) in logarithmic scale. The regions highlighted in blue represent periods of time when the object is observed.

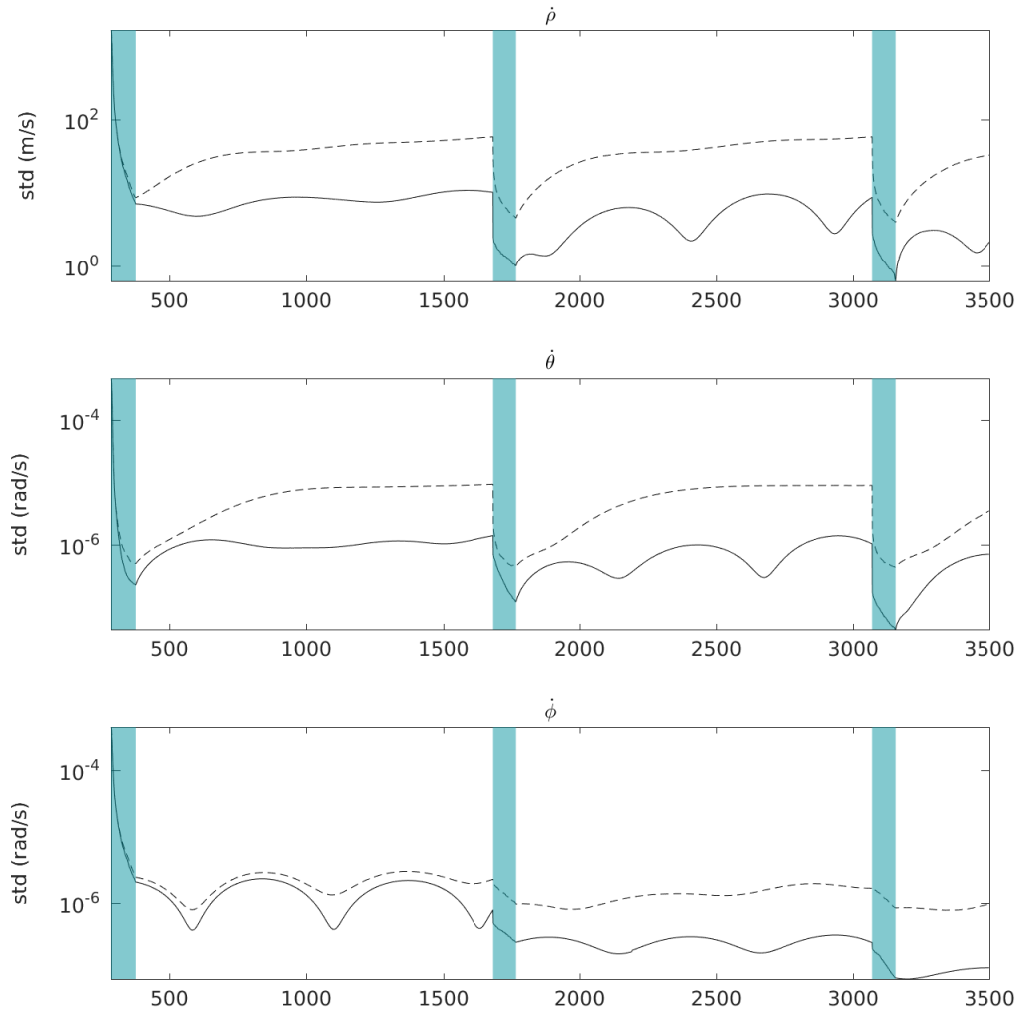


Figure 3.9: Estimated standard deviation for the rates of change of range, azimuth, and elevation, for the orbital elements prediction (continuous line) and the Shepperd matrix prediction (dashed line) in logarithmic scale. The regions highlighted in blue represent periods of time when the object is observed.

experiment as the distance from Earth to the object is not too large. The initialization is done without any prior knowledge of the object's characteristics. Figure 3.10 shows the Root Mean Square Error (RMSE) in position and velocity of the estimate, and the periods when the object is observed. Although the orbit is not perfectly recovered, it can be seen that the estimated uncertainty is sufficient to reacquire the track in the following observation periods and reduce the error each time the object is seen. The estimation error can be seen alongside the estimated parameter variance in figures 3.11 and 3.12, where it can be seen that it is well captured save for a few spikes due to numerical instability.

The second target is on an elliptical orbit which is farther away from Earth. An optical sensor is used to observe the object, which lies in its field of view for most of the filtering duration. The results of the estimation are shown in Figure 3.13, where it can be seen how the error is kept low during the observation period, and starts to drift once the object is no longer seen. As before, the estimated covariance of the filter is shown in figures 3.14 and 3.15, where it can be seen that save for a spike due to numerical errors the estimation error is well captured by the estimated covariance.

3.7 Summary

A method for statistical orbit determination was shown which uses measurements from ground-based radar or optical sensors to provide orbit estimates and their associated uncertainty. The importance of propagating probability distributions rather than point estimates in the space situational awareness context was described. A method was described where particles are used to represent the filtering distributions as this allows for high flexibility in the shape that it can take, which is necessary due to the distribution shapes obtained during initial orbit determination.

The filtering algorithm comprises a method to obtain initial estimates of the probability distribution of the orbit from a single measurement, based on the admissible regions approach. This is done by imposing a constraint on the possible states that the unobserved variables can have, based on the fact that the energy of the observed object must be negative for it to be orbiting the Earth. An orbital propagation method based on the Shepperd matrix was designed, which incorporates uncertainty on the dynamical model to account for unmodeled factors. The update algorithm corrects the distribution when new data is acquired by approximating the predicted particle distribution as a Gaussian distribution in a sensor-centered spherical co-ordinate system, where a linear Kalman update can be applied as the

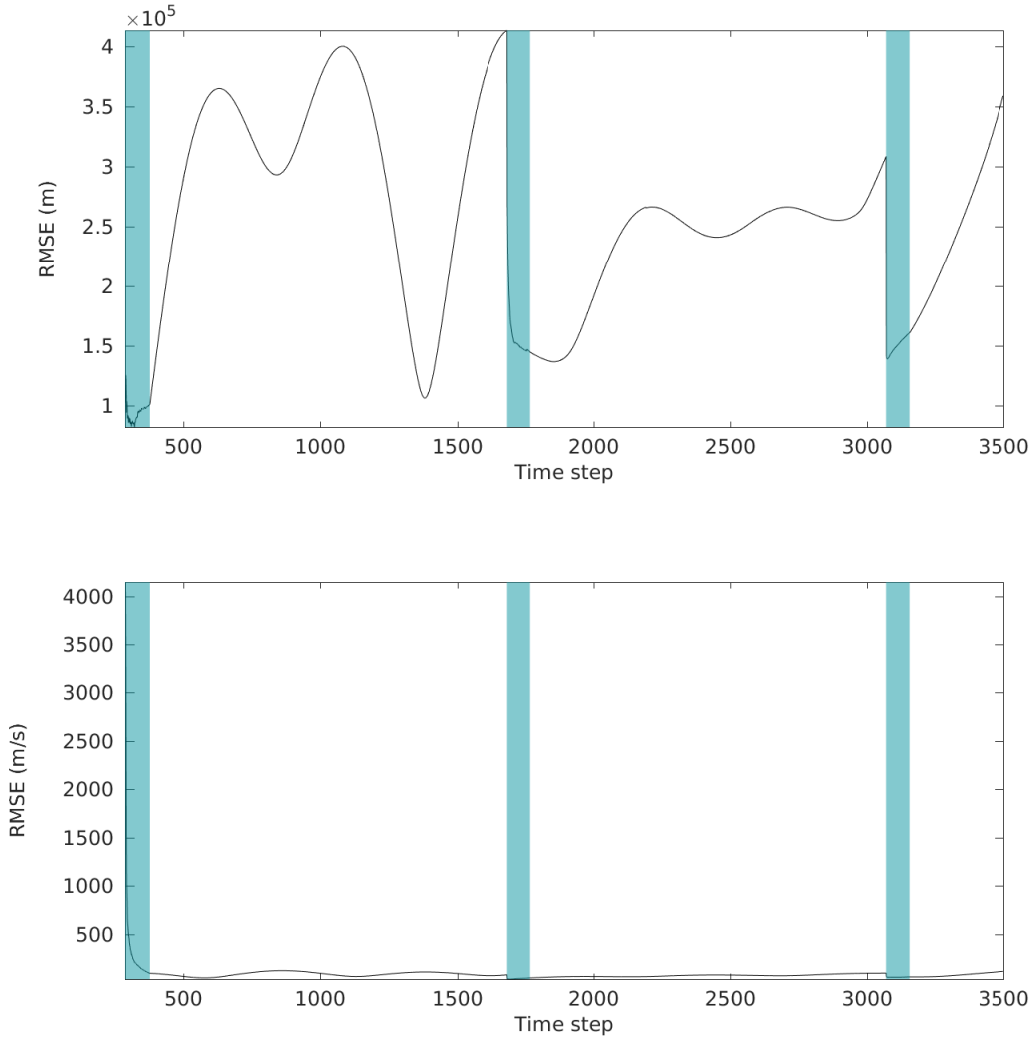


Figure 3.10: Average error in position and velocity for target 1, observed by a radar. The periods when the target is in the field of view are highlighted in blue.

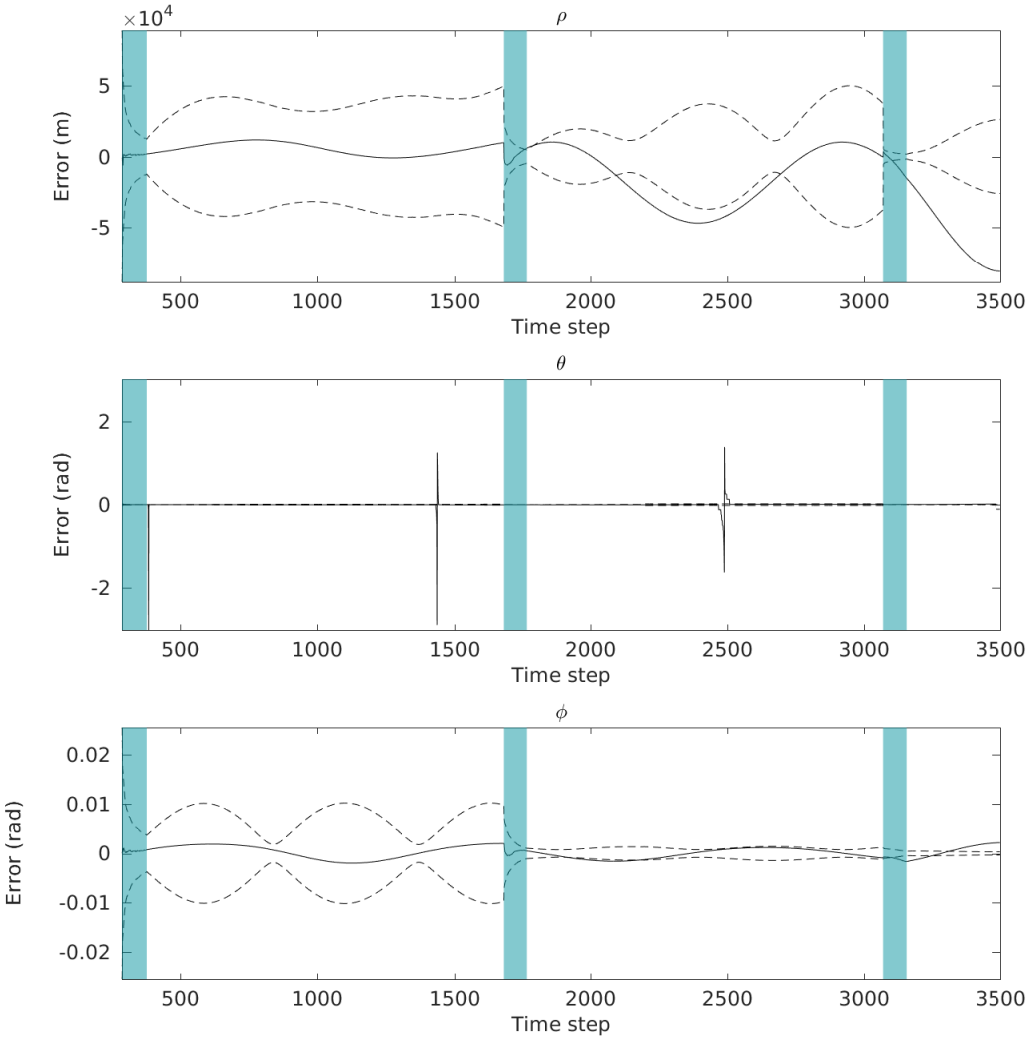


Figure 3.11: Estimated error in spherical co-ordinates (continuous line) and 3-σ bounds of estimated standard deviation (dashed line) for the radar experiment.

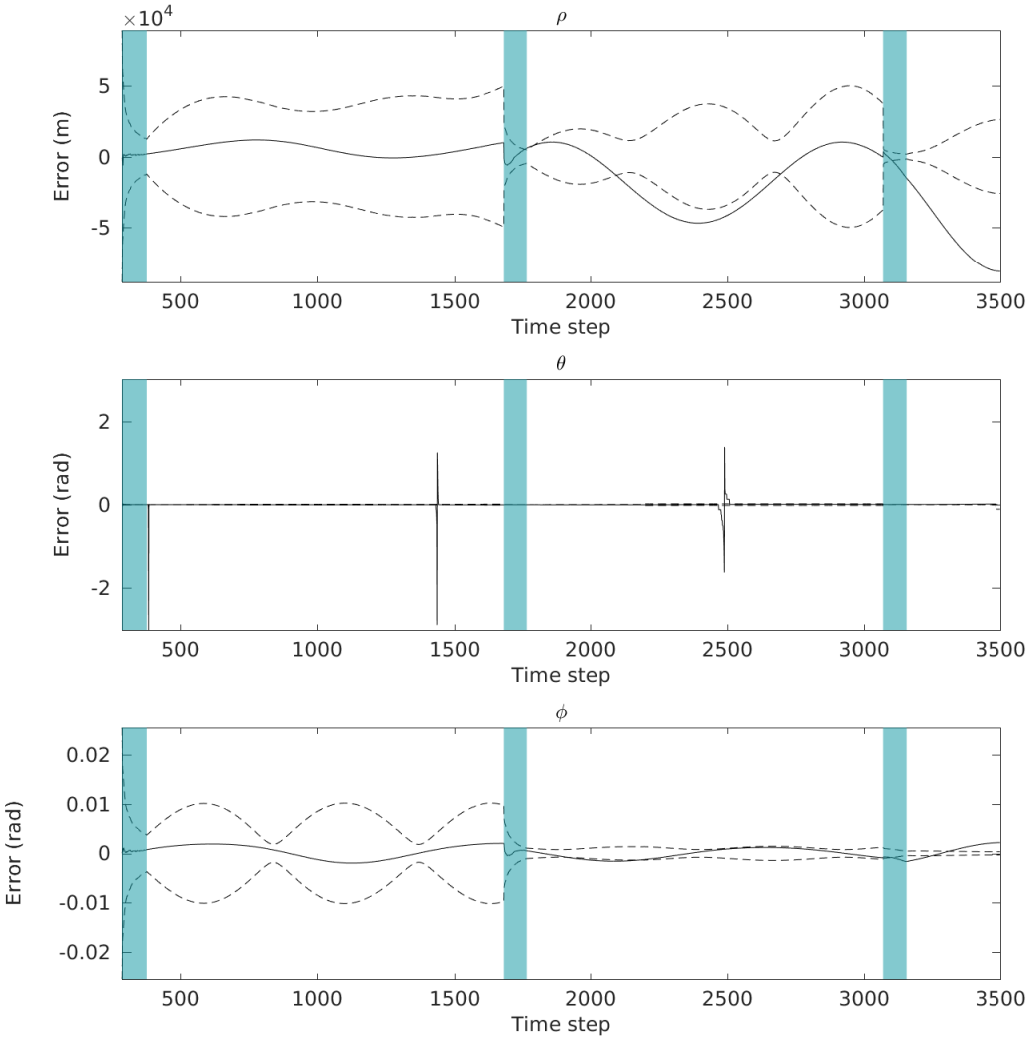


Figure 3.12: Estimated error in spherical co-ordinates (continuous line) and 3-σ bounds of estimated standard deviation (dashed line) for the radar experiment.

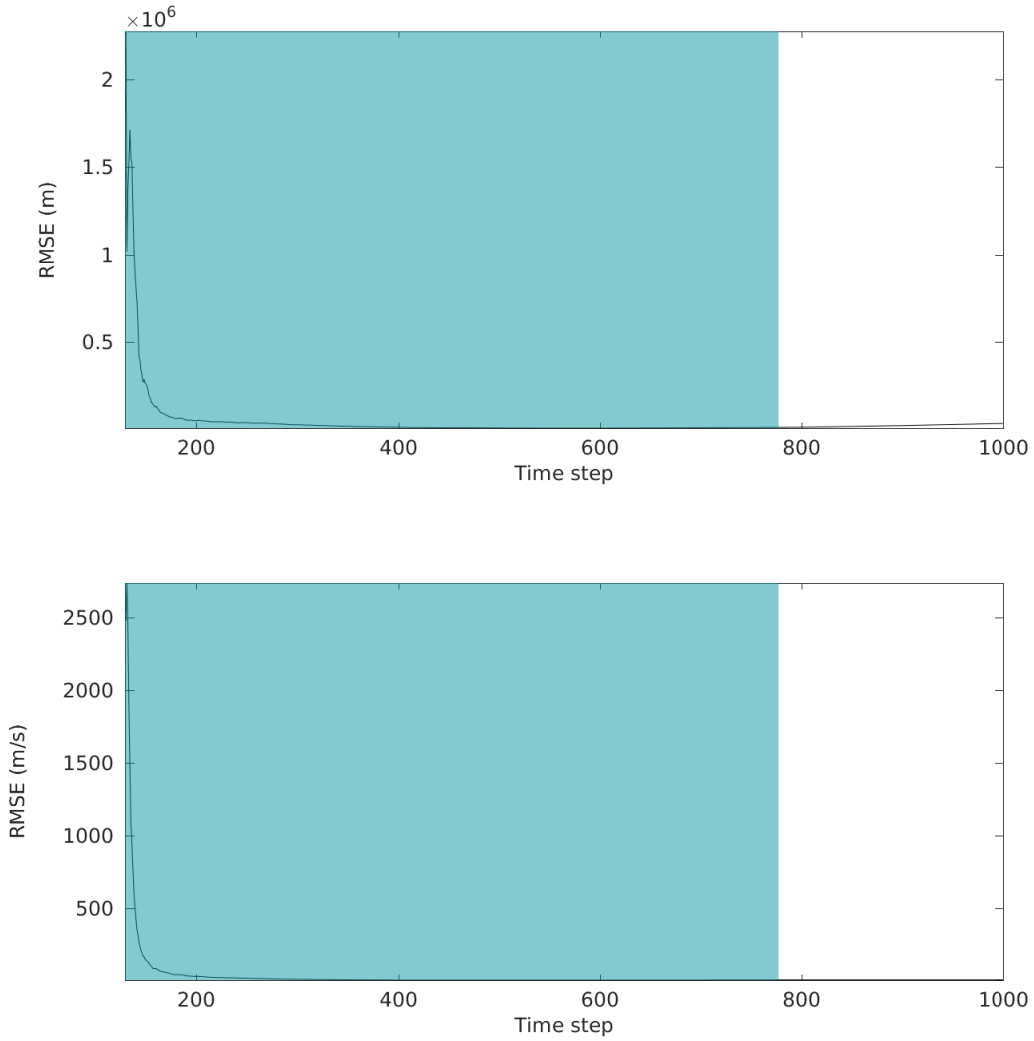


Figure 3.13: Average error in position and velocity for target 2, observed by a telescope. The periods when the target is in the field of view are highlighted in blue.

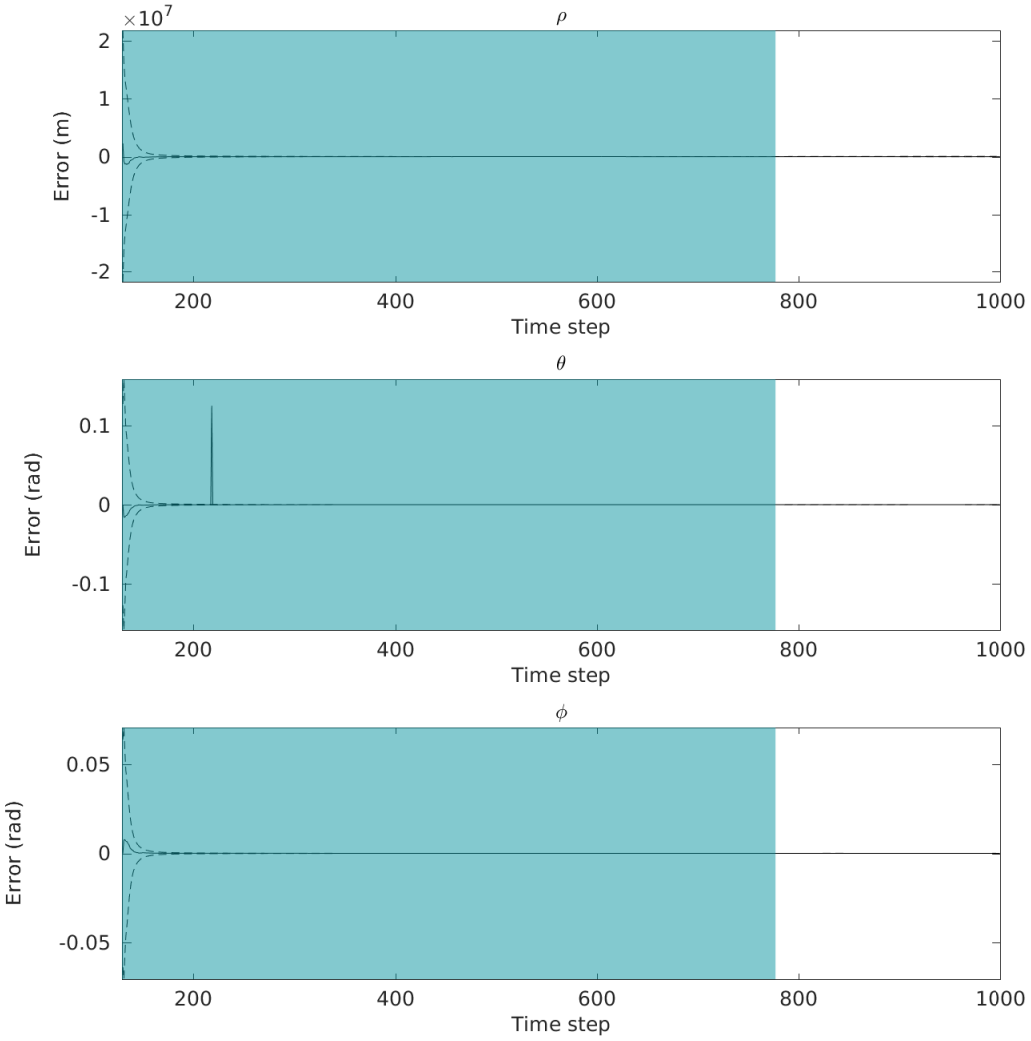


Figure 3.14: Estimated error in spherical co-ordinates (continuous line) and 3-σ bounds of estimated standard deviation (dashed line) for the optical experiment.

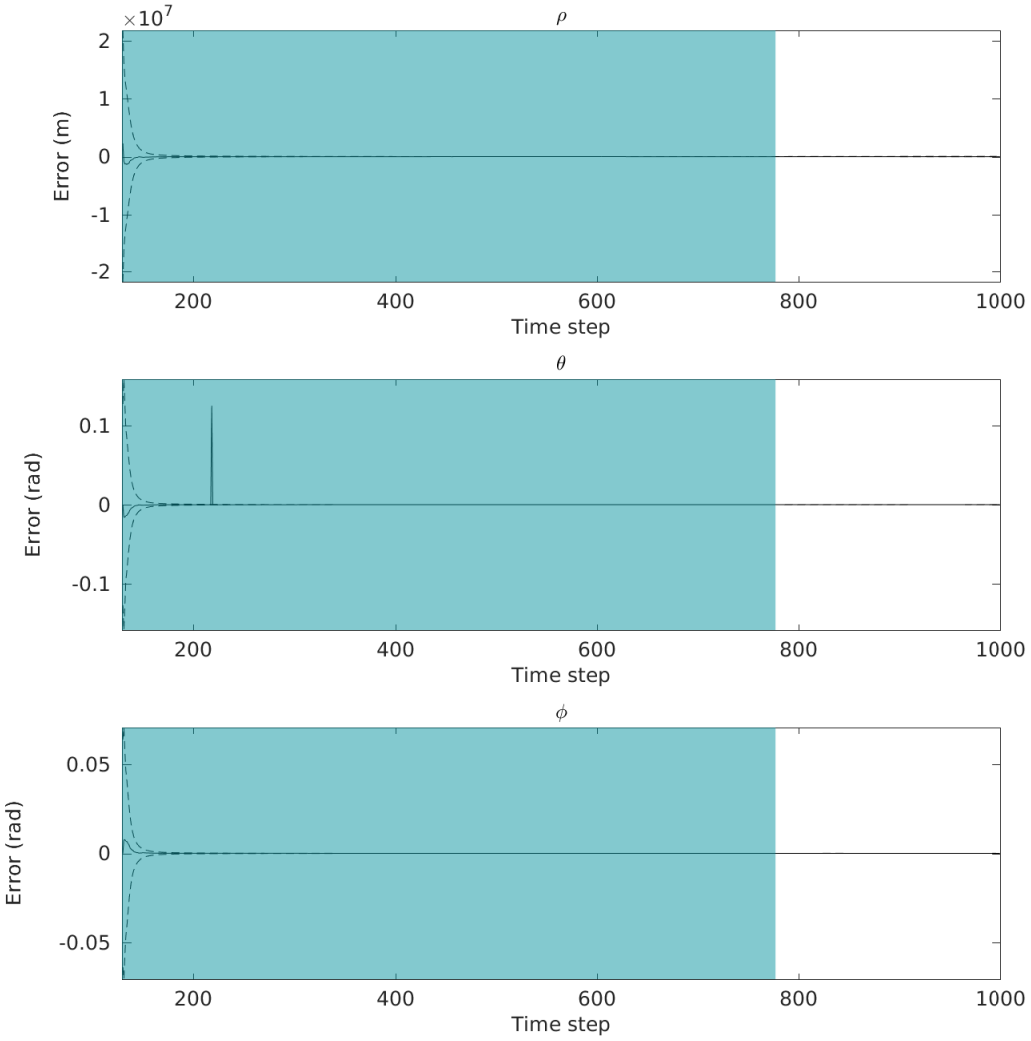


Figure 3.15: Estimated error in spherical co-ordinates (continuous line) and 3-σ bounds of estimated standard deviation (dashed line) for the optical experiment.

measurement model is linear. The validity of this approximation was evaluated using the BHEP test, which tests whether a distribution is Gaussian.

Realistic simulated data was used to validate the approach, and error plots were provided for objects in orbit being observed both with optical and radar sensors. It was shown that for these simulated objects the method has adequate performance. In the next chapter, the approach followed here will be embedded in a multi-object estimation filter in order to estimate populations of objects.

Chapter 4

Efficient State Estimation of Multiple Orbiting Objects

In Chapter 3, the main challenges related to tracking single Earth-orbiting objects were described, and a filter was presented in order to address them. However, the issues related to tracking *populations* of objects were left out. In this chapter, this method is extended to estimate the state of multiple objects in the space situational awareness context.

The maintenance of a catalog of Earth-orbiting objects is essential to preventing collisions, ensuring that an adequate level of accuracy is provided by navigation, communication and surveillance satellites, and identifying potential risks caused by debris. In order to do this, measurements obtained from ground sensors are used to confirm and potentially correct the location of known objects, or identify new ones as they are detected by the sensors. Additionally, it is necessary to consider sensor failings such as missed detections and spurious measurements, and the possibility of ambiguous associations between measurements and tracks.

Tracking objects in the space situational awareness context has several interesting particularities. The field of view of the sensors is small with respect to the surveillance region, which means that targets will not be observed for long periods of time. The dynamics of orbiting objects are well known, which means that the dynamical model will have low amounts of noise. Finally, very little information is available about newly observed objects as they could be debris with very different characteristics, so robust initial orbit determination procedure is required which accurately reflects this lack of knowledge.

A recent estimation framework for stochastic populations proposes an alternative probabilistic description of the multi-object state, by introducing the notion of object

distinguishability [37]. The main feature of this framework is that it allows the representation of both objects that can be individually distinguished, or that are part of a larger population. Two filters have been developed with this framework: The Distinguishable Independent Stochastic Populations (DISP) filter [14], and the Hypothesized Independent Stochastic Populations (HISP) filter [38].

The framework for the estimation of stochastic populations provides the advantages of both random finite set based methods, which provides a way to create principled filters which estimate populations of objects in a fully probabilistic way, with classical approaches which use heuristic track managers to provide individual track information. Another approach which has recently surfaced to do this extends random finite sets with labels to propagate track identity [99], and has produced the labeled multi-Bernoulli (LMB) filter [72]. The LMB filter has cubic complexity, however, whereas the filter that will be proposed in this chapter is linear.

The DISP filter does not make many assumptions apart from the estimated objects behaving independently from one another, and so is highly robust to ambiguous data association. Since it propagates a large amount of information, however, it suffers from high computational complexity. This filter has been successfully used for the estimation of orbiting objects [13], where it was shown that it can deal with high levels of clutter in a multi-sensor scenario.

In space situational awareness, it is often the case that the state of a high number of objects, both objects in the catalog and debris, need to be estimated. Additionally, save for complex scenarios such as the break-up of a satellite, the data association is not usually ambiguous enough to warrant the use of a filter of very high complexity. An alternative filter based on the stochastic populations approach is proposed in this chapter as a way to deal with these issues. The HISP filter [38] is an estimation algorithm that is derived from the DISP by using the assumption that the data association is only moderately ambiguous. The resulting filter loses much of the computational complexity of the DISP, and is linear in the number of targets and in the number of measurements. Due to this, it is ideally suited to a scenario with very large numbers of targets which are observed with sensors with limited coverage, where the measurements are corrupted by noise, and there are possible false alarms and misdetections.

As a base to this multi-object filter, the single-object filter described in Chapter 3 will be employed. An additional improvement that is explored here is a different parametrization of the state space, which permits the use of more accurate orbital dynamics. By using orbital elements [7], the unperturbed two-body problem can be solved directly as opposed to using the linearization that is done when computing

the Shepperd matrix [85]. The effect of using this dynamical model with higher accuracy will be described in this chapter.

The purpose of this chapter is to present a multi-object state estimation filter for space situational awareness which is both powerful and computationally efficient. The framework for estimating stochastic populations is described in section 4.1. The DISP and HISP filters are described in section 4.2, and implementation details of the HISP filter in a space situational awareness scenario are given in section 4.3. Results on a challenging simulated scenario are shown in section 4.4, and finally a summary is presented.

4.1 The estimation framework for stochastic populations

In this section, a broad description of the context in which the HISP filter was developed is presented. The estimation framework for stochastic populations, introduced in [37], proposes a unified probabilistic description of all the sources of uncertainty in a generic multi-sensor multi-object detection and tracking problem. This framework enables the construction of filtering solutions in a principled way, where the choice of the assumptions will determine the complexity of the resulting algorithm. The framework divides the propagated uncertainty in two levels: the individual level and the population level.

On the individual level, the targets of the population of interest are represented by *tracks*. Each track describes the current state of the object, or population of objects, associated to it. A track does not necessarily represent a single target, as it can also represent collectively a sub-population of targets which are indistinguishable from each other for the purpose of estimation. The concept of target distinguishability is a key element of this estimation framework and will be discussed in more detail further on.

On the population level, the composition of the population of interest is represented by multi-target configurations. Each of these proposes a combination of tracks and an associated multiplicity, i.e., the number of targets represented by each track, as a representation of the whole of the population of interest. A multi-target configuration is associated to a scalar weight which describes the probability that this configuration reflects the true composition of the estimated population.

As before, the considered estimation framework is embedded in the Bayesian paradigm. In order to integrate available information about the population of inter-

est, a sequence of time prediction and data update steps are applied to the estimated stochastic population. Time will be indexed by a discrete set $\mathbb{T}$, and for any $t \in \mathbb{T}$, the state space when the target is in the region of interest will be denoted $\boldsymbol{X}_t^{\bullet} \subseteq \mathbb{R}^d$, while the observation space where real measurements are generated will be denoted $\boldsymbol{Z}_t^{\bullet} \subseteq \mathbb{R}^{d'}$. These are augmented with two discrete states; the *empty state* ψ and the *empty observation* ϕ, respectively, to form the *full* target state space $\boldsymbol{X}_t = \boldsymbol{X}_t^{\bullet} \cup \{\psi\}$ and the *full* observation space $\boldsymbol{Z}_t = \boldsymbol{Z}_t^{\bullet} \cup \{\phi\}$. The empty state ψ describes the state of targets outside the region of interest, such as objects that disintregrate upon atmosphere re-entry, or those that leave near-Earth space; while the empty observation ϕ describes misdetections, i.e., when an object fails to produce an observation.

In this application, the target state space $\boldsymbol{X}_t$ is constant throughout the estimation since the surveillance region is a fixed volume of space around Earth. The observation state space $\boldsymbol{Z}_t$, however, can vary since measurements from different sensors can be used for estimation. At each time $t \in \mathbb{T}$, each sensor is assumed to have finite resolution, with resolution cells corresponding to pixels or radar cells, for example. At time $t \in \mathbb{T}$, resolution cells will be indexed by the set Z_t' and the set of observations will be denoted $Z_t \subseteq Z_t'$. When augmented with the empty observation, they will be denoted $\bar{Z}_t = Z_t \cup \{\phi\}$.

4.1.1 Representation of individuals

Each track represents an individual object, or a sub-population of targets that are indistinguishable from one another, as the same information is available for all of them. The probability density p on $\boldsymbol{X}_t$ associated to a track describes the state of each individual of the sub-population of targets that it represents. Examples of populations that are indistinguishable can include, for instance, the pieces of debris following a collision before information on each piece is available to uniquely identify them. In tracking applications, the most direct way to distinguish targets is usually to consider the sequence of measurements which have been associated with it, which will be called their observation history, or observation path. For this reason, The space $\bar{\boldsymbol{O}}_t$ of observation histories is considered, defined as the Cartesian product

$$\bar{\boldsymbol{O}}_t \doteq \bar{Z}_0 \times \cdots \times \bar{Z}_t, \tag{4.1}$$

so that $\boldsymbol{o}_t \in \bar{\boldsymbol{O}}_t$ takes the form $\boldsymbol{o}_t = (\phi, \ldots, \phi, z_{t_+}, \ldots, z_{t_-}, \phi, \ldots, \phi)$ with t_+ and t_- the times of appearance and disappearance of the considered track, respectively, and with $z_t \in \bar{Z}_t$ for any $t \in [t_+, t_-]$. The empty observation path $(\phi, \ldots, \phi) \in \bar{\boldsymbol{O}}_t$ is denoted $\boldsymbol{\phi}_t$. It is worth noting that targets cannot produce measurements before

their time of appearance or after their time of disappearance.

Each target is identified by an index $\boldsymbol{i}$ in a set $\mathbb{I}$. Here, an index refers to any element that uniquely identifies a target, and is not restricted to a number. For instance, observation paths can be a natural way of indexing targets, although even more information can be added to index tuples – for instance, what sub-population a target belongs to, or what time the target disappeared. This will be described in more detail later, as the nature of the index depends on the particular filter.

The DISP and HISP filters rely on the following modeling assumptions. At any time $t \in \mathbb{T}$,

M.1 A target produces at most one observation (if the target does not produce a measurement, a *misdetection* occurs), and

M.2 An observation originates from at most one target (if no target produces it, a *false positive* occurs).

An important consequence of **M.1**, and **M.2** is that an observation characterizes an individual target. A track $\boldsymbol{i}$ associated to an observation path with a least one detection (i.e., $\boldsymbol{o}_t^{\boldsymbol{i}} \neq \boldsymbol{\phi}_t$) cannot have a multiplicity $n^{\boldsymbol{i}}$ greater than one since it cannot represent more than one target. The previously detected target represented by the track $\boldsymbol{i}$ is then said to be distinguishable, and the probability density $p^{\boldsymbol{i}}$ in the full state space $\boldsymbol{X}_t$ describes the state of that individual and none other. On the contrary, a track $\boldsymbol{i}$ associated to the empty observation path $\boldsymbol{o}_t^{\boldsymbol{i}} = \boldsymbol{\phi}_t$ represents a sub-population of yet-to-be-detected targets that are indistinguishable from one another, and may have a multiplicity $n^{\boldsymbol{i}}$ greater than one.

4.1.2 Representation of populations

The composition of the population of interest is described probabilistically using multi-target configurations. These associate a multiplicity to each track in order to describe a specific composition, and attach to this a weight which describes the likelihood of this population given the available data. Whenever the multiplicity of a particular multi-target configuration is zero, it indicates that according to this configuration the target does not exist.

The assumptions that the DISP and HISP filters use that are related to multi-target configurations are the following:

M.3 Targets evolve independently from one another, and

M.4 Measurements that result from target detections are generated independently from one another.

The assumptions of independence are put in place since considering target interactions would make the filters intractably expensive in computational terms. In general, filters derived from this framework propagate tracks covering all possible observation paths from the measurements received up to the current time, along with multi-target configurations representing all possible groups of these tracks and their weights. Additionally, it is possible to propagate information about yet-to-be-detected targets, but in tracking applications this is usually not necessary as little information is available about these.

A simplifying assumption that is used in the HISP filter and some versions of the DISP filter is the following:

S.1 Appearing targets and yet-to-be-detected targets are mixed in a single population

This simplification reduces the complexity of the filters by reducing the amount of sub-populations that need to be propagated. In tracking applications, this does not cause a great loss of precision since it is usually not necessary to estimate the state of objects for which no measurements are available. The population that represents these undetected targets is indexed by the symbol $\mathrm{u} \in \mathbb{I}$.

An illustration of possible observation paths up to time $t = 3$ is shown in Figure 4.1, where two measurements are obtained in time step 1, one in time step 3, and zero in time steps 0 and 2. The six possible observation paths that result from this are listed in the figure.

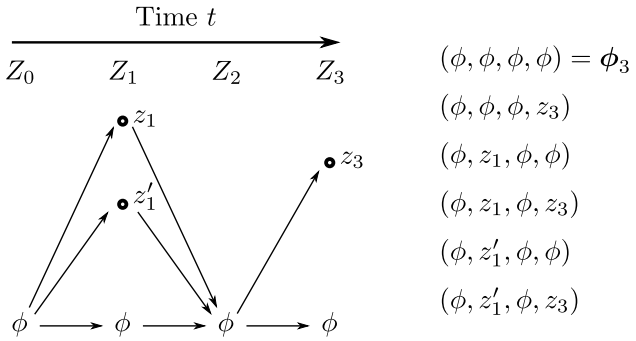


Figure 4.1: Observation paths at time $t = 3$, given a sequence of collected observations.

4.2 Derived filters

In this section, the DISP and HISP filters are described. The DISP filter does not require any additional assumptions apart from the ones described in the previous section, and is a robust estimation algorithm which is well suited to situations where data association is highly ambiguous. The HISP filter uses an additional assumption which discards some information about the population, but benefits from high computational efficiency. Both filters are described below.

4.2.1 The DISP filter

The DISP filter maintains a representation of the stochastic population through a set of tracks indexed through their observation paths $\boldsymbol{i} \in \mathbb{I}$, and a set of hypotheses $\boldsymbol{H}_t$ which are sets of pairwise compatible tracks. In this context, two tracks are said to be pairwise compatible if their observation histories do not share any measurements, which is in line with modeling assumption **M.2**. A possible composition of the population of interest is then given by

- a subset of pairwise compatible tracks $H \subseteq \mathbb{I}_t \setminus \mathrm{u}$ representing tracks that have been previously detected, and
- the undetected track u with multiplicity n^{u} representing a sub-population of n^{u} yet-to-be-detected targets.

The representation of the stochastic population, then, is given by the set of all the possible multi-target configurations $(H, n^{\mathrm{u}}) \in \boldsymbol{H}_t \times \mathbb{N}$ with associated weights $\boldsymbol{w}_t(H, n^{\mathrm{u}}) \in \mathbb{R}^+$ such that

$$\sum_{(H,n^{\mathrm{u}})\in\boldsymbol{H}_t\times\mathbb{N}} \boldsymbol{w}_t(H, n^{\mathrm{u}}) = 1. \tag{4.2}$$

The probability of existence $w_t^{\boldsymbol{i}}$ of distinguishable tracks $\boldsymbol{i} \in \mathbb{I}_t \setminus \mathrm{u}$ can be found by adding the weights of all hypotheses which contain it and all multiplicities of the undetected track:

$$w_t^{\boldsymbol{i}} = \sum_{\substack{(H,n^{\mathrm{u}})\in\boldsymbol{H}_t\times\mathbb{N} \\ H\ni\boldsymbol{i}}} \boldsymbol{w}_t(H, n^{\mathrm{u}}), \quad \boldsymbol{i} \in \mathbb{I}_t. \tag{4.3}$$

Similarly, the cardinality distribution of yet-to-be-detected targets can be found by marginalizing over all hypotheses:

$$\rho_t^{\mathrm{u}}(n) = \sum_{H\in\boldsymbol{H}_t} \boldsymbol{w}_t(H, n), \quad n \in \mathbb{N}. \tag{4.4}$$

The robustness of the DISP filter comes from the fact that it propagates the joint existence of *any* subset of tracks based on their full observation paths. Due to this, its computational complexity grows steeply with the number of observations and targets as the set of maintained hypotheses grows very large. In this sense, it is comparable with the MHT filter, although with the advantage that it is derived from a fully probabilistic framework in a principled way, avoiding the need for track management heuristics.

4.2.2 The HISP filter

The HISP filter is a principled approximation of the DISP filter that considers that it is unlikely for two objects to have generated the same observation in Z_t. In space situational awareness, this is reasonable as orbiting objects tend to have reasonable distances between them with respect to the resolution of the sensors that are used to observe them. This additional assumption greatly simplifies the data update step of the DISP filter, resulting in a method with linear complexity with respect to the number of hypotheses and observations.

In order to describe the prediction and update steps of the HISP filter, it is important to note that as the target state space $\boldsymbol{X}_t = \boldsymbol{X}_t^\bullet \cup \psi$ is a hybrid discrete-continuous space, and so it is necessary to use the formalism of measure theory to define integrals and probabilities. As such, integrals on this space are defined as Lebesgue integrals, such that for instance the integral of a function f on a set $B \subseteq \boldsymbol{X}_t$ is expressed as

$$\int_B f(x)\mathrm{d}x = \int_{\boldsymbol{X}_t^\bullet} \mathbf{1}_B f(x)\mathrm{d}x + \mathbf{1}_B(\psi) f(\psi). \tag{4.5}$$

The observation space is also made up of a continuous part and a discrete part, requiring the use of a special update mechanism. Let p be a probability density function and l_z an integrable function defined on the same space $\boldsymbol{X}_t$. The formula

$$\hat{p}(\boldsymbol{x}) = \frac{l_{\boldsymbol{z}}(\boldsymbol{x})p(\boldsymbol{x})}{\int l_{\boldsymbol{z}}(\boldsymbol{x}')p(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'}, \tag{4.6}$$

defined whenever $\int l_{\boldsymbol{z}}(\boldsymbol{x}')p(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' > 0$, is the equivalent of the Bayes update defined in Chapter 2 with $l_{\boldsymbol{z}}(\boldsymbol{x}) = g(\boldsymbol{z}|\boldsymbol{x})$. The function $l_{\boldsymbol{z}}$ is referred to as a potential since it reshapes p by increasing or decreasing its probability according to the values that it takes. In this filtering application, the potential of interest $l_{\boldsymbol{z}}$ will be the probability density function corresponding to a likelihood on $\boldsymbol{X}_t$ evaluated at the observation

$\boldsymbol{z} \in \boldsymbol{Z}_t^\bullet$.

As it was seen in Chapter 2, if p is Gaussian and the observation process is linear and Gaussian, then $\hat{p}$ is the Kalman filter posterior distribution. In the linear and Gaussian case, solutions to the multi-target tracking problem can be formulated in terms of Kalman filters in interaction [12], and it can be seen that the denominators of the Bayes update, $\int l_{\boldsymbol{z}}(\boldsymbol{x}')p(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'$, indicate the likelihood of track-to-measurement associations.

An issue that comes up in multi-object tracking scenarios is that the use of the measurement likelihood as described in (2.4) can be problematic as its value will be dependent on the reference measure used – the association likelihoods will be different, for instance, if the state space is measured in meters or kilometers, as the association is not unitless (its units are the reciprocal of the units of the reference measure). This means that if different reference measures are used, the value of the association likelihood will be different even if the same association likelihood is being evaluated. A consequence of this is that comparisons to events such as false alarms or newly appearing tracks will be unreliable, as these association functions are dimensionless.

As a way to circumvent this problem, it is preferrable to scale this potential such that the denominator is unitless and is directly comparable to the probabilities of target appearance, survival, and measurement false alarm and misdetection [38], such as

$$l_{\boldsymbol{z}} = \exp\left(-\frac{(H\boldsymbol{m}-\boldsymbol{z})'S^{-1}(H\boldsymbol{m}-\boldsymbol{z})}{2}\right), \tag{4.7}$$

from where

$$\int l_{\boldsymbol{z}}(x)p(x)\mathrm{d}x = \sqrt{\frac{|R|}{|S|}}\exp\left(-\frac{(H\boldsymbol{m}-\boldsymbol{z})'S^{-1}(H\boldsymbol{m}-\boldsymbol{z})}{2}\right), \tag{4.8}$$

where $|\cdot|$ denotes the determinant. This potential is unitless, scales the association likelihoods to the range $(0, 1]$, and does not change the results of the single-target estimated probability as the scaling affects the numerator and denominator equally.

In practice, the observation process at time t is modeled by a potential

$$\begin{aligned} \ell_t^{\boldsymbol{z}}(\boldsymbol{x}) &= p_{\mathrm{d,t}}(\boldsymbol{x})l_t^{\boldsymbol{z}}(\boldsymbol{x}), \quad \boldsymbol{z} \in Z_t, \\ \ell_t^{\phi}(\boldsymbol{x}) &= 1 - p_{\mathrm{d,t}}(\boldsymbol{x}), \end{aligned} \tag{4.9}$$

where $p_{\mathrm{d,t}}(\boldsymbol{x})$ is the probability of detection and the dimensionless potential $l_t^{\boldsymbol{z}}$ is the likelihood of association with measurement $\boldsymbol{z}$.

Simplification **S.1** implies that in terms of population management, the HISP filter collapses the populations of *appearing targets*, i.e., those entering the surveillance region at the current time step; and *yet-to-be-detected targets*, i.e., those that have previously entered the scene but have not yet been detected, into the same track. Due to this, the time of appearance of a target is unknown and does not provide a way to uniquely identify an individual in the population. Since targets with the same observation paths can both represent objects that disappeared from the scene at the last time they were observed, and objects that have been misdetected since the last measurement, it is important to add the last time at which the target is believed to have been in the scene to the indexing function.

At time step $t \in \mathbb{T}$, targets are then indexed by pairs $\boldsymbol{i} = (t', \boldsymbol{o}) \in \mathbb{I}_t = \mathbb{T} \times \bar{\mathbb{O}}_t$ where t' is the last epoch where the target was known to be in the scene, and the observation path $\boldsymbol{o}$ indicates the measurements that have been associated with it. If $t' < t$, it means that the target is believed to have left the scene, whereas if $t' = t$ the target is believed to still be in the scene. Individuals of the first type are not used for filtering, but they are used for state extraction as will be explained later, so they will be stored in a different indexing set, $\mathbb{I}_t^-$ for targets which disappeared at time t, which will be defined later.

The composition of a population after prediction at time t is denoted $\mathbb{I}_{t|t-1} = \{(t, \boldsymbol{o} \,|\, \boldsymbol{o} \in \bar{\mathbb{O}}_{t-1}\}$, while the updated population is indexed with set $\mathbb{I}_t = \{(t, \boldsymbol{o} \,|\, \boldsymbol{o} \in \bar{\mathbb{O}}_t\}$. From here on, the symbol (m) will be used to denote tracks that have been previously detected (or measured), such that $\boldsymbol{o} \neq \boldsymbol{\phi}_t$. This is the population of previously-detected targets. The symbol (u) will be used to denote targets that are in the state space, but have not yet been detected, in which case $\boldsymbol{o} = \boldsymbol{\phi}_t$. According to **S.1**, this last population has a single element, which will be denoted $\boldsymbol{i}_t^{\mathrm{u}}$, or when there is no possible ambiguity, simply u. Using this notation, the HISP filter can be expressed by the propagation of a set of hypotheses

$$\mathcal{P}_t = \left\{ \left(p_t^{\boldsymbol{i}}, w_t^{\boldsymbol{i}}, n_t^{\boldsymbol{i}} \right) \right\}_{\boldsymbol{i} \in \mathbb{I}_t} .$$

For each $\boldsymbol{i} \in \mathbb{I}_t$, $p_t^{\boldsymbol{i}}$ indicates the single-object probability distribution associated to the track, $w_t^{\boldsymbol{i}}$ is the weight or probability of existence, and $n_t^{\boldsymbol{i}}$ is the multiplicity of the hypothesis. It is important to note that the probability of existence of a track is different from the probability of presence in the scene, $1 - p_t^{\boldsymbol{i}}(\psi)$. Also, the HISP filter propagates hypotheses that are different in meaning than those of the DISP, the biggest difference being that a single set is propagated with all the tracks. Track extraction is then more involved than the DISP, where it all that is required is to

extract the set H with the highest weight.

In order to maintain a tractable structure for the HISP filter, some additional assumptions need to be made. In particular, it is assumed that the terms

$$\breve{w}_t^{\boldsymbol{k},\boldsymbol{z}} = w_{t|t-1}^{\boldsymbol{k}} \int \ell_t^{\boldsymbol{z}}(\boldsymbol{x}) p_{t|t-1}^{\boldsymbol{k}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{4.10}$$

corresponding to the association of target with index $\boldsymbol{k} \in \mathbb{I}_{t|t-1}$ with observation $\boldsymbol{z} \in Z_t$ satisfy, for tracks $\boldsymbol{k}, \boldsymbol{k}' \in \mathbb{I}_{t|t-1}$, that it is unlikely for two tracks to produce the same measurement:

S.2 For any $\boldsymbol{k}, \boldsymbol{k}' \in \mathbb{I}_{t|t-1}$ with $\boldsymbol{k} \neq \boldsymbol{k}'$, and any $\boldsymbol{z} \in Z_t$, it holds that $\breve{w}_t^{\boldsymbol{k},\boldsymbol{z}} \breve{w}_t^{\boldsymbol{k}',\boldsymbol{z}} \approx 0$.

The final assumption sacrifices the propagation of the joint probability of existence of tracks, and is also necessary to obtain linear computational complexity:

S.3 Hypotheses are independent of one another.

HISP prediction is done through a transition kernel which is divided in three parts: one which models object dynamics, one which models object disappearance and one which models object appearance. The first kernel models exclusively the dynamics of the object, and is denoted q_t^{π}. For $\boldsymbol{x} \in \boldsymbol{X}_t^{\bullet}$ and $\boldsymbol{x}' \in \boldsymbol{X}_{t-1}^{\bullet}$, it is the transition kernel described in Chapter 2: $q_t^{\pi}(\boldsymbol{x}, \boldsymbol{x}') = f(\boldsymbol{x}|\boldsymbol{x}')$. For $\boldsymbol{x}' \in \boldsymbol{X}_{t-1}^{\bullet}$, $q_t^{\pi}(\psi, \psi) = 1$ and $q_t^{\pi}(\boldsymbol{x}', \psi) = 0$. The probability of survival can be computed as $p_t^{\pi}(\boldsymbol{x}) = \int q_t^{\pi}(\boldsymbol{x}, \boldsymbol{x}') \mathrm{d}\boldsymbol{x}'$ and can be interpreted as the probability that a target with state $\boldsymbol{x}$ does not disappear at time $t-1$.

The disappearance kernel is denoted q_t^{ω} and satisfies, for $\boldsymbol{x}' \in \boldsymbol{X}_{t-1}^{\bullet}$, $\int_{\boldsymbol{X}_t^{\bullet}} q_t^{\omega}(\boldsymbol{x}', \boldsymbol{x}) = 1$ and $q_t^{\omega}(\psi, \boldsymbol{x}) = 0$. The transitions q_t^{π} and q_t^{ω} are meant to be complementary in the sense that $q_t^{\omega}(\boldsymbol{x}, \psi) + p_t^{\pi}(\boldsymbol{x})$ – either a target disappears or it does not. The final kernel q_t^{α} is for appearing targets; for $\boldsymbol{x}' \in \boldsymbol{X}_{t-1}^{\bullet}$ and $\boldsymbol{x} \in \boldsymbol{X}_t^{\bullet}$, $q_t^{\alpha}(\boldsymbol{x}', \boldsymbol{x}) = 0$ and $q_t^{\alpha}(\boldsymbol{x}', \psi) = 0$, and $q_t^{\alpha}(\psi, \boldsymbol{x})$ is the distribution of appearing targets with weight w_t^{α}.

After prediction, the newly appeared and yet-to-be-detected targets are jointly represented by

$$p_{t|t-1}^{\mathrm{u}}(\boldsymbol{x}) = \frac{n_{t-1}^{\mathrm{u}} \int q_t^{\pi}(\boldsymbol{x}', \boldsymbol{x}) p_{t-1}^{\mathrm{u}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}' + n_t^{\alpha} p_t^{\alpha}(\boldsymbol{x})}{n_{t-1}^{\mathrm{u}} + n_t^{\alpha}} \tag{4.11}$$

$$w_{t|t-1}^{\mathrm{u}} = \frac{n_{t-1}^{\mathrm{u}} w_{t-1}^{\mathrm{u}} + n_t^{\alpha} w_t^{\alpha}}{n_{t-1}^{\mathrm{u}} + n_t^{\alpha}} \tag{4.12}$$

$$n_{t|t-1}^{\mathrm{u}} = n_{t-1}^{\mathrm{u}} + n_t^{\alpha}, \tag{4.13}$$

where $(p_t^{\alpha}, w_t^{\alpha}, n_t^{\alpha})$ are the probability distribution, weight, and cardinality of appearing targets at time t.

The remainder of the predicted population is composed of objects that have been observed at least once in the past, and have indices of the form $\boldsymbol{k} = (t-1, \boldsymbol{o})$ with $\boldsymbol{o} \neq \boldsymbol{\phi}_{t-1}$. These can either be propagated using kernel q_t^{π}, or disappear with kernel q_t^{ω}. After prediction, they are represented by

$$p_{t|t-1}^{\boldsymbol{i}}(\boldsymbol{x}) = \int q_t^{\kappa}(\boldsymbol{x}', \boldsymbol{x}) p_{t-1}^{\boldsymbol{k}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}', \tag{4.14}$$

$$w_{t|t-1}^{\boldsymbol{i}} = (w_{t|t-1}^{i}, 1), \tag{4.15}$$

$$n_{t|t-1}^{\boldsymbol{i}} = 1 \tag{4.16}$$

for $\kappa \in \{\pi, \omega\}$, and $\boldsymbol{i} = (t, \boldsymbol{o})$ when $\kappa = \pi$ and $(t-1, \boldsymbol{o})$ otherwise. These last hypotheses are not considered further for filtering – they are not indexed in $\mathbb{I}_{t|t-1}$ – but must be stored in the set of disappeared targets $\mathbb{I}_t^-$ for the track extraction process. The resulting multi-target configuration is then defined as

$$\mathcal{P}_{t|t-1} = \left\{ \left(p_{t|t-1}^{\boldsymbol{i}}, w_{t|t-1}^{\boldsymbol{i}}, n_{t|t-1}^{\boldsymbol{i}} \right) \right\}_{\boldsymbol{i} \in \mathbb{I}_{t|t-1}}$$

.

For the observation update, it is necessary to model false alarms and misdetections. The probability that a false alarm will be generated in a resolution cell $\boldsymbol{z} \in Z_t$ will be denoted $v_t^{\boldsymbol{z}}$. For $\boldsymbol{k} = (t, \boldsymbol{o}) \in \mathbb{I}_{t|t-1}$ and $\boldsymbol{z} \in \bar{Z}_t$ (recall $\bar{Z}_t$ is the set of resolution cells augmented with the empty detection), define index $\boldsymbol{i} = (t, \boldsymbol{o} \times \boldsymbol{z})$ where this observation path is the concatenation of $\boldsymbol{o}$ and measurement $\boldsymbol{z}$. Let $p_t^{\boldsymbol{i}}$ be the probability density function on $\boldsymbol{X}_t$ given by

$$p_t^{\boldsymbol{i}}(\boldsymbol{x}) = \frac{\ell_t^{\boldsymbol{z}} p_{t|t-1}^{\boldsymbol{k}}(\boldsymbol{x})}{\int \ell_t^{\boldsymbol{z}} p_{t|t-1}^{\boldsymbol{k}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}'}. \tag{4.17}$$

In order to obtain the posterior probability for $\boldsymbol{z} \in Z_t$ to be a false alarm, and to help with state extraction, an additional set of indices needs to be introduced to represent false alarms. At time step t, let $\mathbb{I}_t' = \{\boldsymbol{z}\}_{\boldsymbol{z} \in Z_t}$, and define $w_t^{\boldsymbol{z},\boldsymbol{z}} = \breve{w}_t^{\boldsymbol{z},\boldsymbol{z}} = v_t^{\boldsymbol{z}}$, and $w_t^{\boldsymbol{z},\phi} = 1 - v_t^{\boldsymbol{z}}$. Then for $\boldsymbol{k} \in \mathbb{I}_{t|t-1} \cup \mathbb{I}_t'$, the weights corresponding to the *a posteriori* probability for the given hypothesis are given by

$$w_t^{\boldsymbol{i}} = \frac{w_{\mathrm{ex}}^{\boldsymbol{k},\boldsymbol{z}} \breve{w}_t^{\boldsymbol{k},\boldsymbol{z}}}{\sum_{\boldsymbol{z}' \in \bar{Z}_t} w_{\mathrm{ex}}^{\boldsymbol{k},\boldsymbol{z}'} w_t^{\boldsymbol{k},\boldsymbol{z}'}} \tag{4.18}$$

or, equivalently,

$$w_t^{\boldsymbol{i}} = \frac{w_{\text{ex}}^{\boldsymbol{k},\boldsymbol{z}} \breve{w}_t^{\boldsymbol{k},\boldsymbol{z}}}{\sum_{\boldsymbol{k}' \in \mathbb{I}_{t|t-1}} w_{\text{ex}}^{\boldsymbol{k}',\boldsymbol{z}} w_t^{\boldsymbol{k}',\boldsymbol{z}}}, \tag{4.19}$$

where $w_t^{\boldsymbol{k},\boldsymbol{z}} = \breve{w}_t^{\boldsymbol{k},\boldsymbol{z}} + \mathbf{1}_{\boldsymbol{\phi}}(\boldsymbol{z})(1 - w_{t|t-1}^{\boldsymbol{k}})$ is the probability mass attributed to the association between hypothesis with index $\boldsymbol{k}$ and measurement $\boldsymbol{z}$, which also accommodates the possibility of the target not existing in the case $\boldsymbol{z} = \phi$, i.e., a missed detection. For an observation $\boldsymbol{z} \in Z_t$ and an index $\boldsymbol{k} \in \mathbb{I}_{t|t-1} \cup \mathbb{I}'_t$, the scalar $w_{\text{ex}}^{\boldsymbol{k},\boldsymbol{z}}$ is the probability of the association of the remaining observations with false alarms, any of the remaining hypotheses, or any of the remaining yet-to-be-detected individuals. Approximation **S.2** permits this to be expressed as

$$w_{\text{ex}}^{\boldsymbol{k},\boldsymbol{z}} = C'_t(\boldsymbol{k}, \boldsymbol{z}) \prod_{\boldsymbol{k}' \in \mathbb{I}^{\text{m}}_{t|t-1} \setminus \{\boldsymbol{k}\}} \left[w_t^{\boldsymbol{k}',\phi} + \sum_{\boldsymbol{z}' \in Z_t \setminus \{\boldsymbol{z}\}} \frac{w_t^{\boldsymbol{k}',\boldsymbol{z}'}}{C_t(\boldsymbol{z}')} \right], \tag{4.20}$$

where $\mathbb{I}^{\text{m}}_{t|t-1} = \mathbb{I}_{t|t-1} \setminus \boldsymbol{i}^{\text{u}}_{t-1}$, where $C_t(\boldsymbol{z}) = w_t^{\text{u},\boldsymbol{z}}/w_t^{\text{u},\phi} + v_t^{\boldsymbol{z}}/(1 - v_t^{\boldsymbol{z}})$ and where

$$C'_t(\boldsymbol{k}, \boldsymbol{z}) = \left[w_t^{\text{u},\phi}\right]^{n^{\text{u}}_{t|t-1} - \mathbf{1}_{\text{u}}(\boldsymbol{k})} \left[\prod_{\boldsymbol{z}' \in Z'_t \setminus Z'} (1 - v_t^{\boldsymbol{z}'}) \right] \left[\prod_{\boldsymbol{z}' \in Z_t \setminus \{\boldsymbol{z}\}} C_t(\boldsymbol{z}') \right] \tag{4.21}$$

with $Z' = \{\boldsymbol{z}\}$ when $\boldsymbol{k} \in \mathbb{I}'_t$ and $Z' = \emptyset$ otherwise. After the update, the resulting state configuration is given by

$$\mathcal{P}_t = \left\{ \left(p_t^{\boldsymbol{i}}, w_t^{\boldsymbol{i}}, n_t^{\boldsymbol{i}}\right) \right\}_{\boldsymbol{i} \in \mathbb{I}_t},$$

where $n_t^{\boldsymbol{i}} = n^{\text{u}}_{t|t-1}$ if $\boldsymbol{i} = \text{u}$ and $n_t^{\boldsymbol{i}} = 1$ otherwise.

As it can be seen, the computation of target weights is of linear complexity with respect to the number of measurements and the number of hypotheses. Assumptions **S.2** and **S.3** are essential to this, and a derivation of the filter using them can be seen in [37].

4.3 The HISP filter for space situational awareness

This section describes the implementation of the HISP filter in a space situational awareness scenario, extending the single object estimation method described in section 3 to track multiple orbiting targets.

4.3.1 State representation and dynamical model

As it was seen in chapter 3, the orbital elements parametrization gives the possibility to model noise in such a way that the predicted covariance of the filter does not grow too fast, without adding such little noise that it becomes hard to associate measurements and tracks. For this reason, it will be used in this chapter.

Although it is possible to model target disappearance (for example, objects disintegrating as they enter the atmosphere of the Earth), it is out of the scope of this work. The disappearance kernel q_t^{ω} is assumed constant and verifies, for $x \in \boldsymbol{X}_t^{\bullet}$, $q_t^{\omega}(x, \psi) = 10^{-10}$, that is, the probability of survival is assumed to be almost unity.

4.3.2 Initial orbit determination and data update

The initial orbit determination method that is used is the same as the one described in Chapter 3. The appearance kernel q_t^{α} is also assumed to be constant, as no *a priori* information about target appearance is available.

The data update step is modified in order to consider the field of view of the sensor. As opposed to the single-object filter described in Chapter 3, it is necessary to introduce particle weights in order to consider the probability of detection. In this application, it is determined to be a sensor-dependent constant if the state is within the field of view of the sensor, and zero elsewhere. If track $\boldsymbol{i} \in \mathbb{I}_{t|t-1}$ is associated with the empty measurement ϕ, then the particle weights $\gamma_{t-1}^{\boldsymbol{k},j}, j = 1 \dots J$, are updated as

$$\gamma_t^{\boldsymbol{i},j} = \frac{\left[1 - p_{\mathrm{d},t}(\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j})\right]\gamma_{t-1}^{\boldsymbol{k},j}}{\sum_{j'=1}^{J}\left[1 - p_{\mathrm{d},t}(\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j'})\right]\gamma_{t-1}^{\boldsymbol{k},j'}}, \tag{4.22}$$

where $\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j}$ is the state of the j-th particle of track $\boldsymbol{k}$. This gives an association weight

$$\breve{w}_t^{\boldsymbol{k},\phi} = w_{t|t-1}^{\boldsymbol{k}} \sum_{j=1}^{J}\left[1 - p_{\mathrm{d,t}}(\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j})\right]\gamma_{t-1}^{\boldsymbol{k},j}. \tag{4.23}$$

In the case that $\boldsymbol{z} \neq \phi$, the target is detected and needs to be updated. The state update is in this case the same as in Chapter 3, save that the particles are now weighted, and the probability of detection needs to be taken into account. The Gaussian approximation is obtained by first weighting the predicted distribution $p_{t|t-1}^{\boldsymbol{k}}$ with the probability of detection, and obtaining the Gaussian approximation

in spherical co-ordinates with these weighted samples:

$$\begin{aligned}
\hat{\gamma}_t^{\boldsymbol{k},j} &= \frac{\gamma_t^{\boldsymbol{k},j} p_{\mathrm{d},t}(\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j})}{\sum_{j'=1}^{J} p_{\mathrm{d},t}(\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j'}) \gamma_t^{\boldsymbol{k},j'}}, \quad j = 1, \ldots, J \\
\boldsymbol{\mu}_{t|t-1}^{\boldsymbol{k}} &= \sum_{j=1}^{J} \hat{\gamma}_t^{\boldsymbol{k},j} \boldsymbol{y}_{t|t-1}^{\boldsymbol{k},j} \\
P_{t|t-1}^{\boldsymbol{k}} &= \sum_{j=1}^{J} \hat{\gamma}_t^{\boldsymbol{k},j} \left(\boldsymbol{y}_{t|t-1}^{\boldsymbol{k},j} - \boldsymbol{\mu}_{t|t-1}^{\boldsymbol{k}} \right) \left(\boldsymbol{y}_{t|t-1}^{\boldsymbol{k},j} - \boldsymbol{\mu}_{t|t-1}^{\boldsymbol{k}} \right)'
\end{aligned} \tag{4.24}$$

where $\boldsymbol{y}_{t|t-1}^{\boldsymbol{k},j}$ is obtained by transforming $\boldsymbol{x}_{t|t-1}^{\boldsymbol{k},j}$ into spherical co-ordinates as described in Chapter 3. As before, a Kalman update is applied in this space which permits the computation of the association weight as in the linear and Gaussian case:

$$\breve{w}_t^{\boldsymbol{k},\boldsymbol{z}} = w_{t|t-1}^{\boldsymbol{k}} \left(\sum_{j=1}^{J} \hat{\gamma}_t^{\boldsymbol{k},j} \right) \sqrt{\frac{|R|}{\left| S_{t|t-1}^{\boldsymbol{k}} \right|}} \exp\left(\frac{1}{2} \left(H \boldsymbol{\mu}_{t|t-1}^{\boldsymbol{k}} - \boldsymbol{z} \right)' \left(S_{t|t-1}^{\boldsymbol{k}} \right)^{-1} \left(H \boldsymbol{\mu}_{t|t-1}^{\boldsymbol{k}} - \boldsymbol{z} \right) \right) \tag{4.25}$$

4.3.3 State extraction

One of the necessary tradeoffs of the HISP filter is that it loses information on the joint probability of existence of targets. As such, in order to extract the most plausible configuration from the propagated population $\mathcal{P}_t$, it is necessary to perform an additional step. A criterion to extract the likeliest configuration is to obtain a subset of tracks which have the maximum possible a posteriori weight, while maintaining pairwise compatibility. Additionally, it is important to consider the likelihood of measurements being false alarms, weighted against the likelihood of association (which has been stored in set $\mathbb{I}_t'$), and the disappeared targets in a given window τ prior to the current time step. The set of indices that are used in the optimization process is then

$$\mathbb{I}_{\mathrm{ext}} = \mathbb{I}_t \cup \bigcup_{k=t-\tau}^{t} \left(\mathbb{I}_k' \cup \mathbb{I}_k^- \right), \tag{4.26}$$

and the desired subset I_t of targets can be found by solving

$$I_t = \underset{I \subseteq \mathbb{I}_{\mathrm{ext}}}{\operatorname{argmax}} \prod_{\boldsymbol{i} \in I} w_t^{\boldsymbol{i}} \tag{4.27}$$

subject to the constraint of no targets sharing any measurements in their observation paths. This is equivalent to solving

$$I_t = \underset{I \subseteq \mathbb{I}_{\text{ext}}}{\operatorname{argmax}} \sum_{\boldsymbol{i} \in I} \ln w_t^{\boldsymbol{i}} \tag{4.28}$$

subject to the same constraint, which can be done using integer programming [101].

4.3.4 Additional approximations

With the process described in this section, the set of assumptions grows at every iteration. This is undesirable as it will mean that obtaining the posteriors at each time step will always grow in computational cost, and all hypotheses, even those that are very unlikely, will be propagated.

As an additional simplifying assumption, it can be considered that certain hypotheses grow too unlikely to be of interest to the filter. As such, after each time step, hypotheses $\boldsymbol{i} \in \mathbb{I}_t$ such that $w_t^{\boldsymbol{i}} < \tau$, with τ a user-defined threshold that indicates the minimum likelihood for which tracks are still of interest, can be removed from the indexing set without a high loss of information.

4.4 Results

To evaluate the performance of the HISP filter, simulated data was generating with a perturbed orbital model considering the Earth's gravitational field up to degree and order 12; third order perturbations of the Sun and the Moon; and direct radiation pressure. The propagation was done with a Runge-Kutta numerical integration procedurre, where 30 spherical-shaped satellites were simulated.

In order to observe the objects, three sweeping Doppler radar sensors were simulated, located at latitudes 15°, 0°, and −15°, with an even spread in longitude. The sensors sweep along the parallel of latitude on a 120° arc with an angular speed of $\frac{2\pi}{5000}\text{rad}\,\text{s}^{-1}$. The sensor resolution and the profile of the field of view of all sensors can be seen in Table 4.1, and a constant probability of false alarm was chosen on the volume of the sensor with mean 1 false alarm per time-step per sensor. The false alarm rate, obtained by dividing this among the number of resolution cells of the sensor, is $v_t^{\boldsymbol{z}} = 7.7 \times 10^{-15}$. The probability of detection was chosen to be constant in the field of view and equal to $p_{\text{d}} = 0.98$.

The orbital parameters of the simulated objects can be seen in Table 4.2, and a snapshot of their positions in orbit can be seen in Figure 4.3, along with the field of

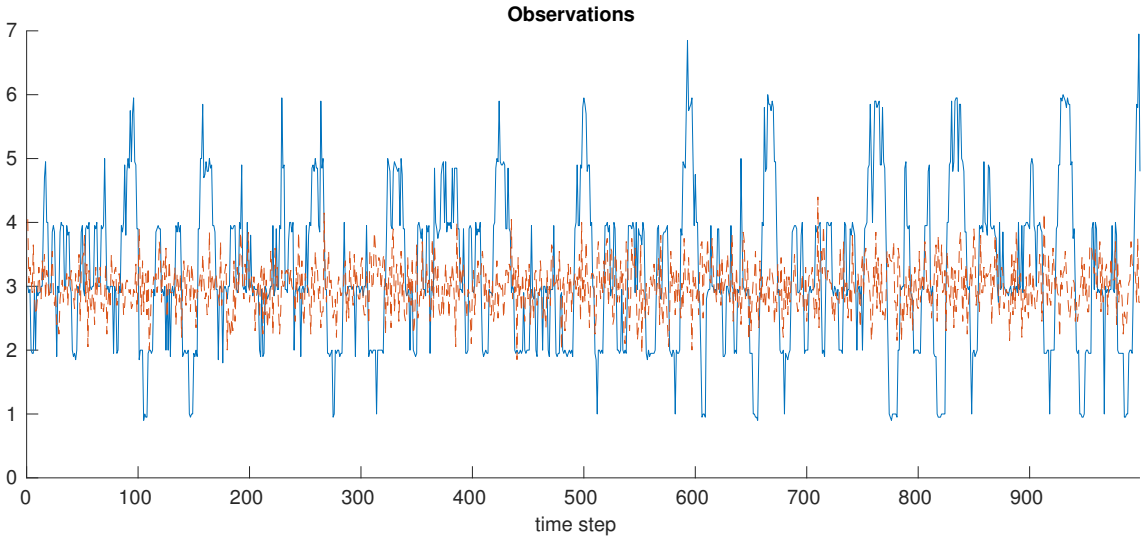


Figure 4.2: Amount of observations per time step in the simulated scenario. True measurements in blue, and clutter in red.

	r (m)	θ (°)	φ (°)	$\dot{r}$ ($\mathrm{m\,s^{-1}}$)
Cell resolution	100	0.1	0.1	10
Noise (std. dev.)	100	0.1	0.1	10
Field of view	$[50,\ 45 \times 10^6]$	$[-8,\ 8]$	$[-45,\ 45]$	$[-1 \times 10^4,\ 1 \times 10^4]$

Table 4.1: Sensor resolution and field of view profile: range r, azimuth θ, elevation φ, range rate $\dot{r}$.

view of the sensors. Figure 4.2 shows the measurements per time step observed by all of the sensor, separating the true positives and the false alarms.

In terms of modeling, the matrix Q_t was defined to be constant across the scenario and equal to

$$Q_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^{-10} \end{bmatrix}, \tag{4.29}$$

and the matrix R_t was also chosen to be a constant matrix with entries equal to those found in Table 4.1 in the diagonal, squared, and zeros elsewhere. The particle number was set to $J = 500$, and a pruning threshold of 10^{-4} was used to curtail the growth of the number of hypotheses.

The scenario was run for 1000 time steps with a constant time lapse of $\Delta_t =$

ID	a (km)	e	i (°)	Ω (°)	ω (°)	ν (°)
1	42170.23	0.000973	35.74	359.30	124.11	341.97
2	42190.79	0.000492	2.64	295.41	255.24	317.11
3	42164.04	0.000724	2.44	295.04	203.94	90.98
4	42153.98	0.000899	4.64	302.39	221.78	44.81
5	42292.17	0.000292	14.44	339.69	146.81	352.69
6	42164.64	0.000141	13.88	100.70	16.65	163.95
7	42212.84	0.000733	14.93	10.88	178.06	306.36
8	42359.18	0.000915	15.02	14.21	94.00	179.74
9	42249.35	0.000779	14.17	27.57	160.13	102.09
10	42308.74	0.000319	13.81	30.98	346.16	386.31
11	42415.51	0.000439	11.59	41.16	175.70	247.43
12	24520.92	0.718	6.54	166.94	338.94	193.98
13	42525.07	0.000697	10.94	42.45	109.67	302.51
14	24937.89	0.710	63.00	227.43	288.57	199.96
15	42165.47	0.000381	63.00	72.85	158.97	127.09
16	42165.90	0.000333	0.0258	109.85	105.54	227.60
17	42165.04	0.000341	0.0247	174.29	347.57	254.69
18	24727.66	0.716	1.00	106.26	59.20	146.90
19	42165.57	0.000470	0.0215	242.27	354.11	65.60
20	42166.58	0.000146	0.0240	214.79	359.35	282.24
21	42166.67	0.000166	0.0515	123.26	79.71	301.46
22	42164.80	0.00157	7.43	52.77	114.58	6.10
23	42165.97	0.000279	0.0283	89.19	131.19	96.03
24	42165.37	0.000269	0.0332	236.95	21.52	28.76
25	24738.09	0.721	88.19	313.00	287.64	164.39
26	24637.14	0.687	64.85	186.72	272.84	205.14
27	42170.58	0.00345	54.43	203.14	195.94	106.66
28	42166.40	0.000668	2.72	245.36	327.12	342.87
29	42165.96	0.000378	0.0310	31.02	188.23	81.98
30	24353.42	0.725	17.85	272.76	181.96	148.44

Table 4.2: Initial semi-major axis a, eccentricity e, inclination i, right ascension of the ascending node Ω, argument of perigee ω, and true anomaly ν of the 30 orbiting objects in the scenario.

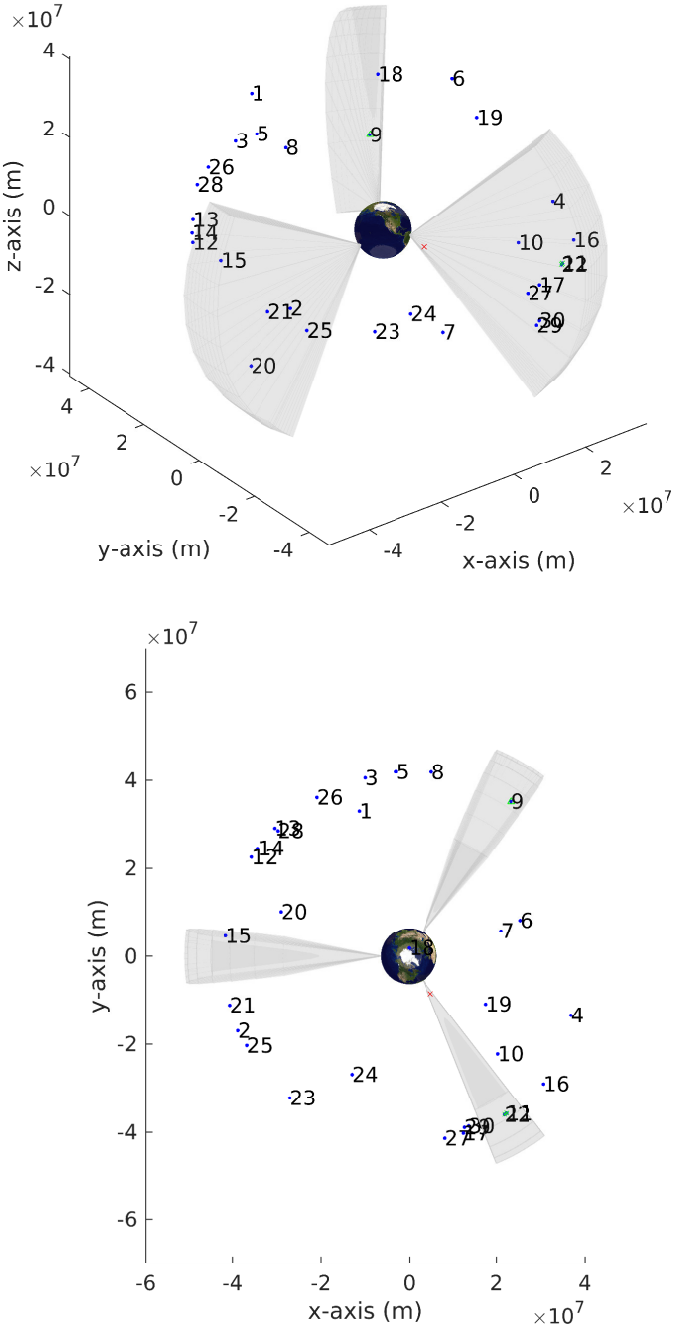


Figure 4.3: Near-Earth space with the initial position of the 30 targets (blue dots) and the sensor field of view (grey volumes), in 3D view (above) and 2D view from above (below).

20s between time steps. In Figure 4.4, the estimated number of objects is shown alongside the observable number of objects. An object is said to be observable from the first time step it enters the field of view of any sensor. Here it can be seen that given the sharpness of the used models, and the low amount of false alarms, the HISP is highly reactive and takes about two time steps to initialize tracks. Two spikes are visible where the HISP filter briefly overestimated the number of targets.

Two sample tracks were selected to illustrate the single-object estimation performance, by showing the estimated number of tracks originated by measurements produced by the chosen object; the number of track swaps, or events when the filter decides that a track with a different observation path is more credible than the one it had been building up up to that time step (it "changes its mind" about the track it is following); and the root mean square errors for position and velocity for the best track in case the number of tracks is overestimated.

Figures 4.5 and 4.6 show the estimation results for object 1. Here, the HISP filter performed very well in the estimation. In Figure 4.5 it can be seen that the number of tracks is not overestimated at any time in the scenario, and that the track was initialized very quickly after it was observed for the first time. There are no track swaps, which indicates that the filter was confident about the observation path that it assigned to this object. Figure 4.6 shows how the RMSE of position and velocity decreases steadily as the object is observed, and the estimation improves as it is reobserved. All objects save for objects 4 and 12 show behavior similar to this one.

Figures 4.7 and 4.8 show the performance of the filter in estimating the state of track 12. Figure 4.7 shows that the HISP filter discards the track which it had assigned to the object in favor of restarting with the most recent measurements. This can be explained by the fact that the dynamical model was not accurate enough to correctly predict the behavior of the target, and the introduced bias caused the filter to assign more credibility to a freshly started track for it. The resulting RMSE, which can be seen in Figure 4.8, is accordingly higher across the scenario. The estimated state for object 4 shows similar behavior.

It can be noted that for both shown tracks the error seems to rise slightly during the observation process when the obtained error is close to the lowest available error. This can be due to several reasons - the estimator for the mean position that is used to compute the RMSE is relatively crude, as it was discussed in Chapter 3. Additionally, the particle resampling that is carried out every time an update is done can also add an amount of approximation error. Although in general the filter accurately reduces the estimation error of the observed objects, these observed spikes open the path to investigate how to preserve particle histories, which could smooth

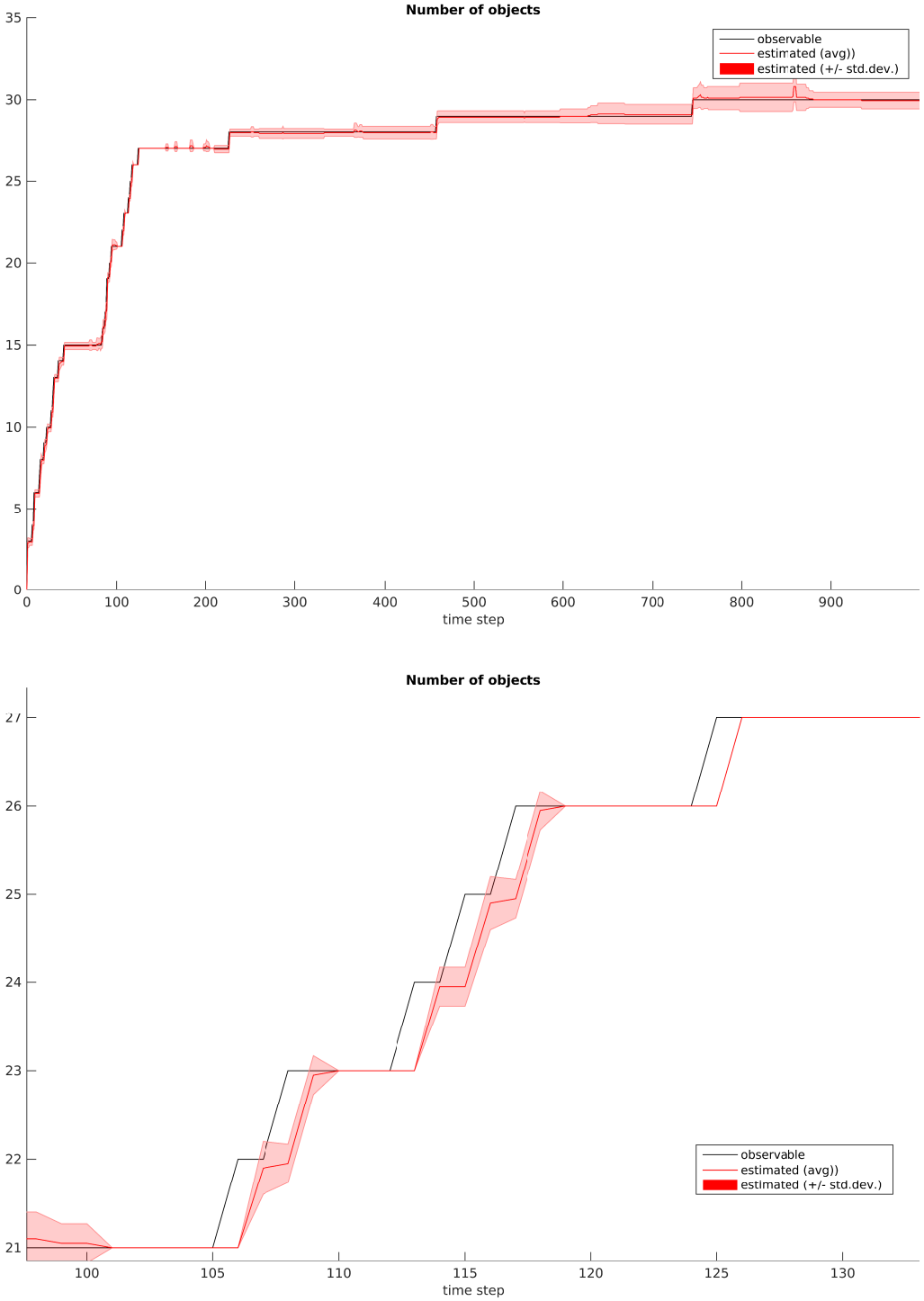


Figure 4.4: Cardinality estimate, full scenario (above), detail (below). An object is deemed observable from its first time of entry in one of the sensor field of view.

these spikes over as the amount of resampling done would be greatly reduced.

Figure 4.9 shows the run time across the scenario, where the interesting thing to note is that the trend of the plot follows the number of observable objects, which is in line with the linear complexity of the filter in terms of tracks and measurements. This is due to the number of hypotheses maintained by the filter being reasonably low. In Figure 4.10, the amount of hypotheses per timestep of the filter can be seen in a representative run, where it is clear that the number of hypotheses maintained is kept to a reasonable number.

4.5 Summary

In this chapter the HISP filter, a novel multi-object estimation algorithm, is presented in the context of space situational awareness. This filter originates from a recent estimation framework for stochastic populations, and is an approximation of the DISP filter which had previously been used to estimate a smaller number of orbiting targets.

The presented filter enables the extension of the single-object estimation algorithm presented in Chapter 3 to track a number of orbiting objects, and account for the problems inherent to multiple object estimation. The complexity of the algorithm is linear in the number of objects and the number of measurements which is ideal for situations where a large number of objects need to be estimated, such as catalog maintenance in space situational awareness.

As an improvement to the work presented in Chapter 3, a more accurate dynamical model is presented which exploits the use of orbital elements in order to solve the unperturbed two-body problem.

The filter was tested with simulated data, showing good performance. It is reactive in the creation of new tracks, while remaining robust to false alarms and misdetections and performing its computations in reasonable amounts of time. Future work includes implementing a parallel version of the filter, as many computations can be done independently, and improving the dynamical model further to better account for perturbations.

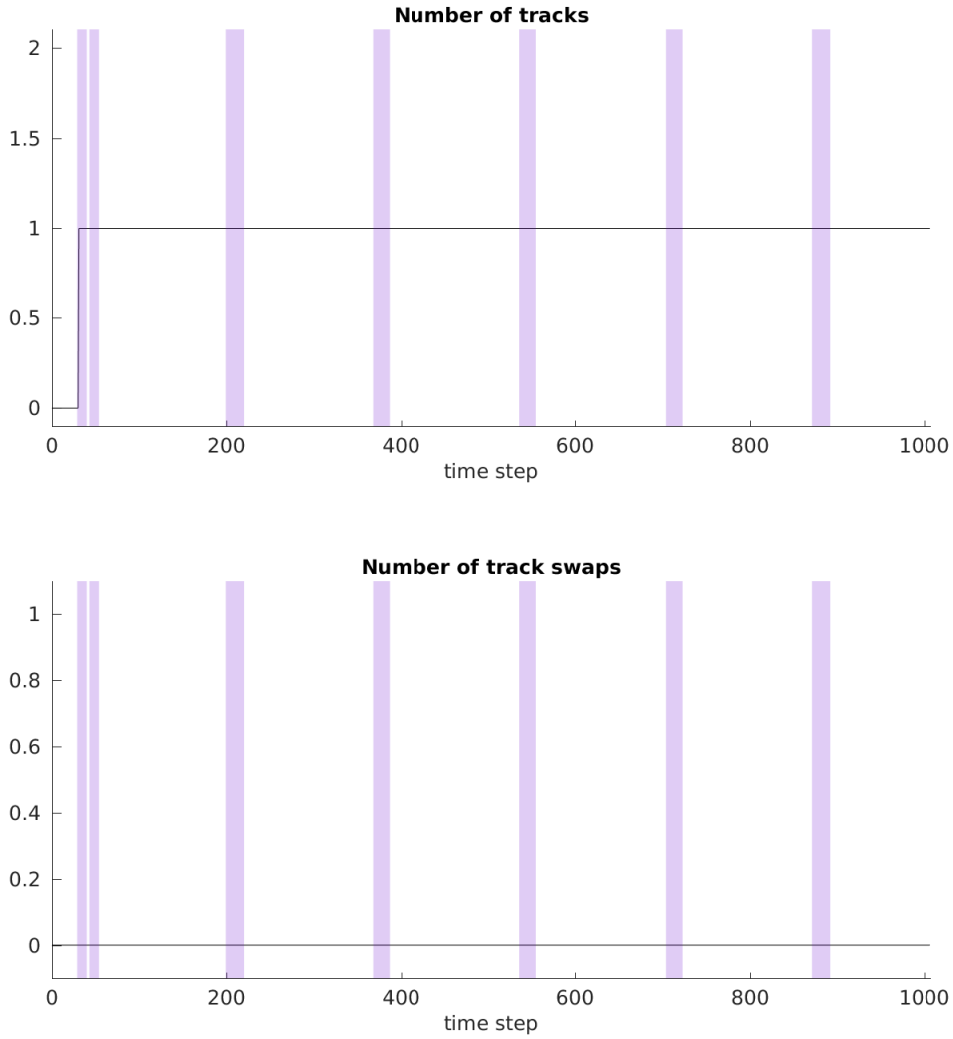


Figure 4.5: Estimated number of tracks and track swaps for object 1. Blue areas indicate time windows where the object is in the field of view of a sensor.

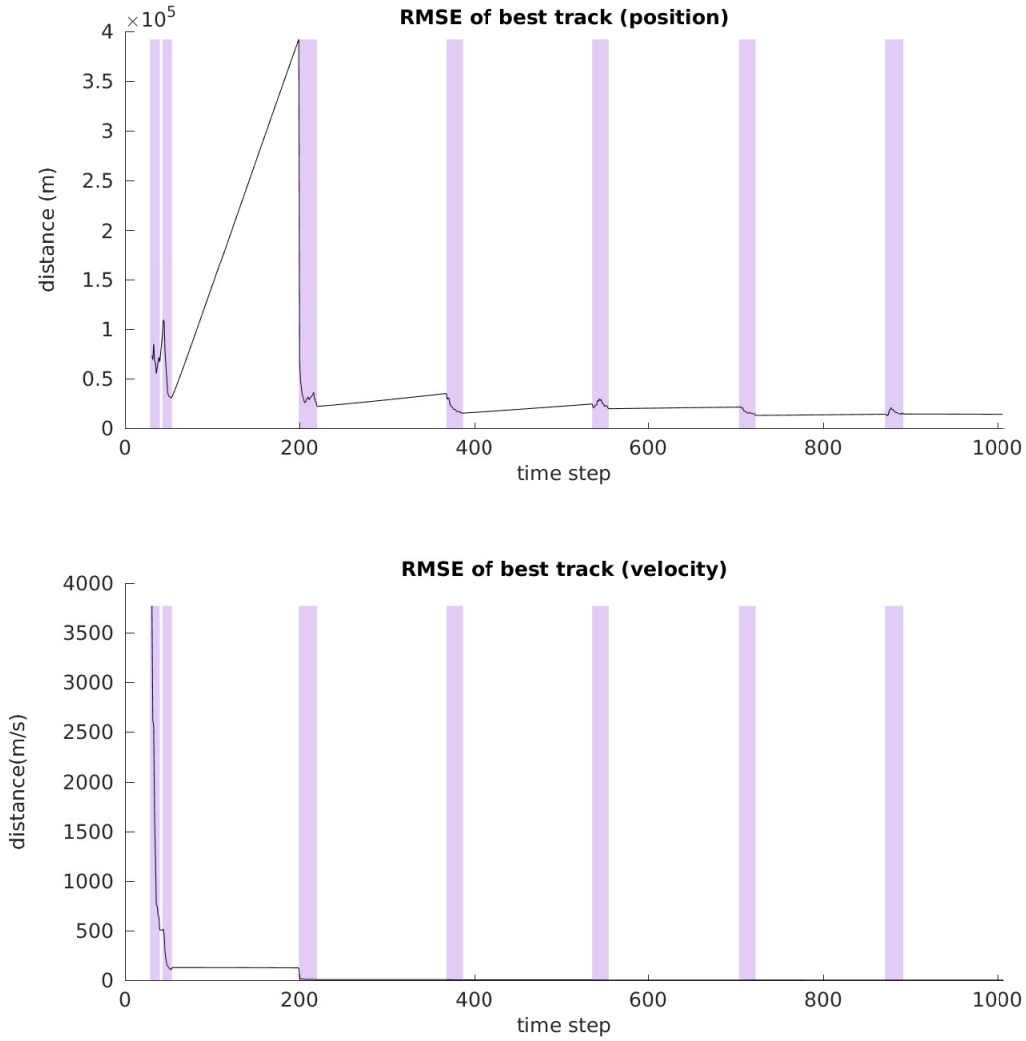


Figure 4.6: Position and velocity RMSE for object 1. Blue areas indicate time windows where the object is in the field of view of a sensor.

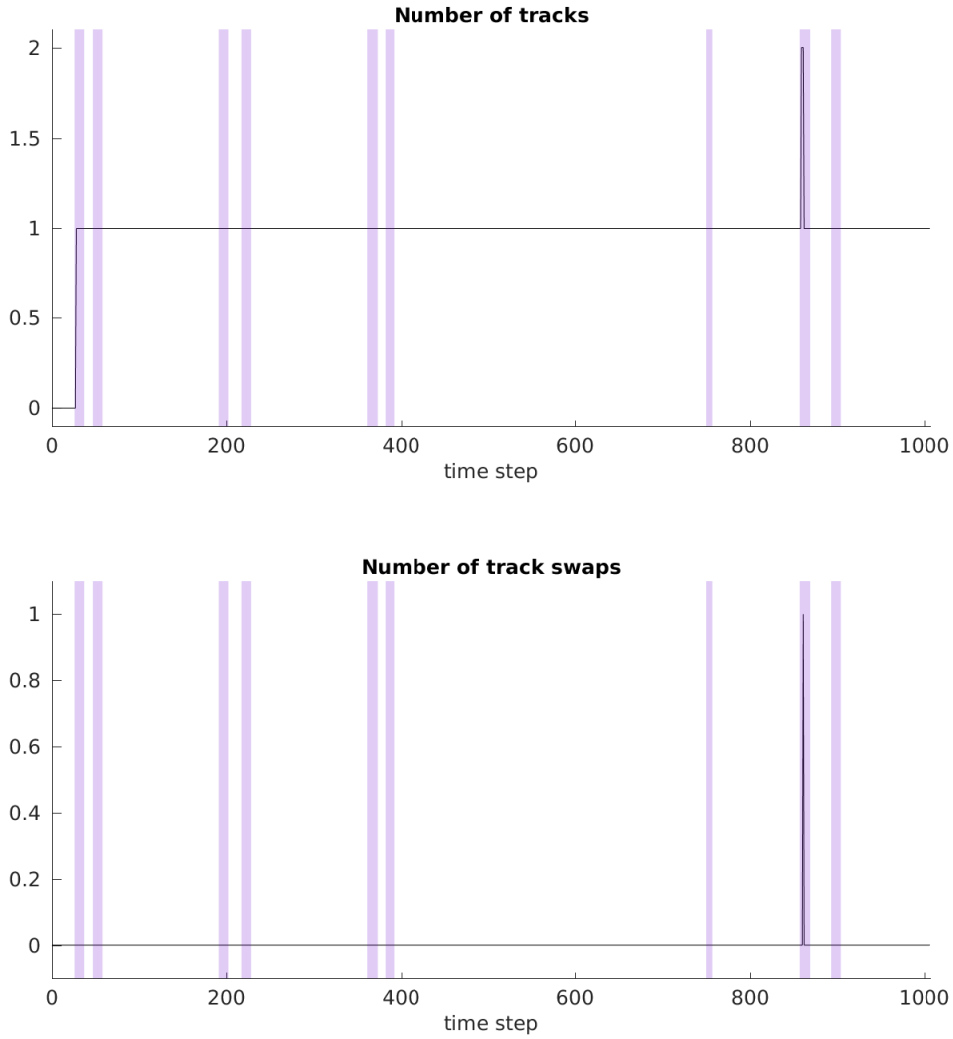


Figure 4.7: Estimated number of tracks and track swaps for object 12. Blue areas indicate time windows where the object is in the field of view of a sensor.

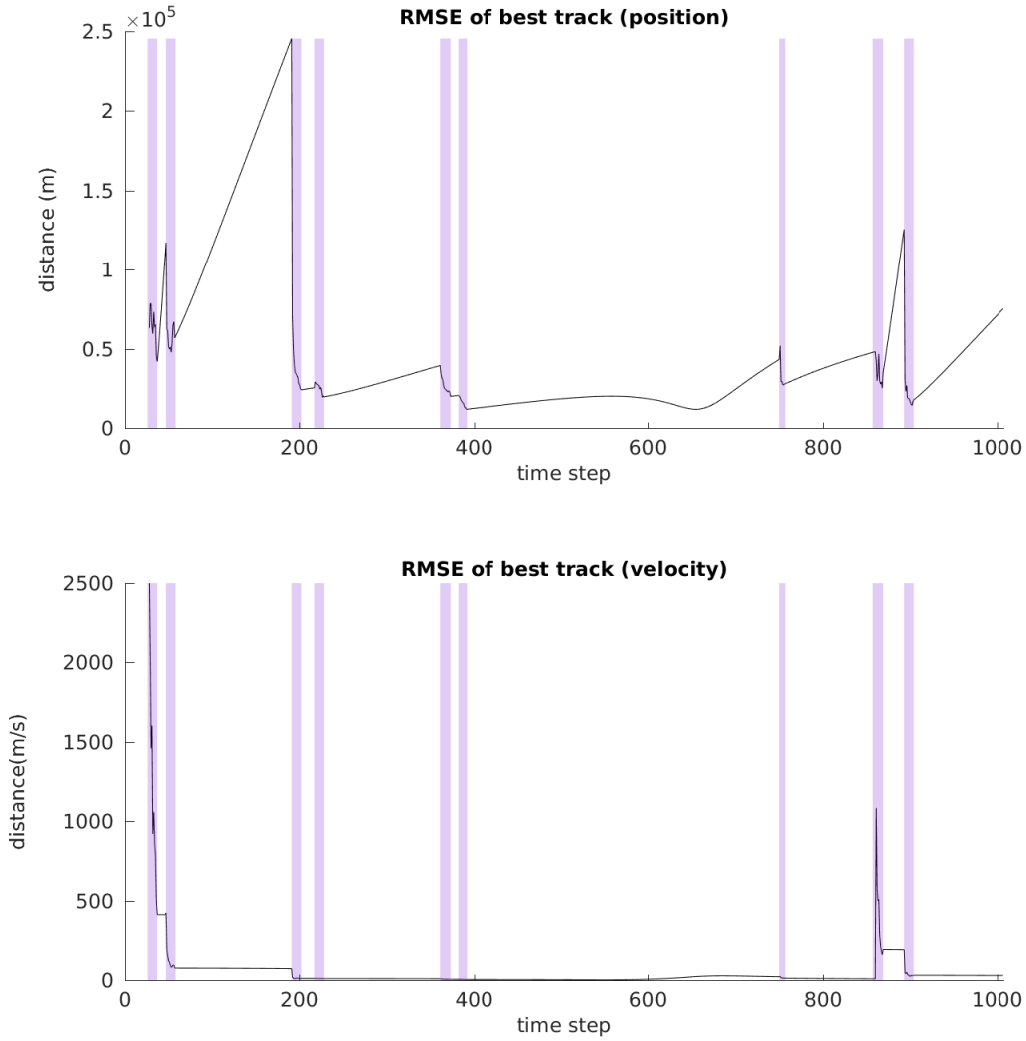


Figure 4.8: Position and velocity RMSE for object 12. Blue areas indicate time windows where the object is in the field of view of a sensor.

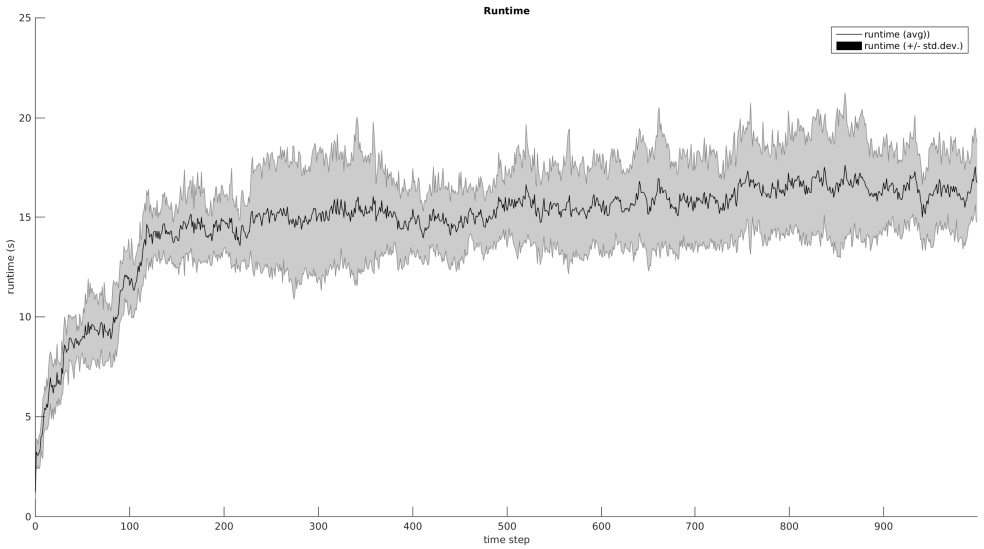


Figure 4.9: Processing time for each time step of the filter across the scenario.

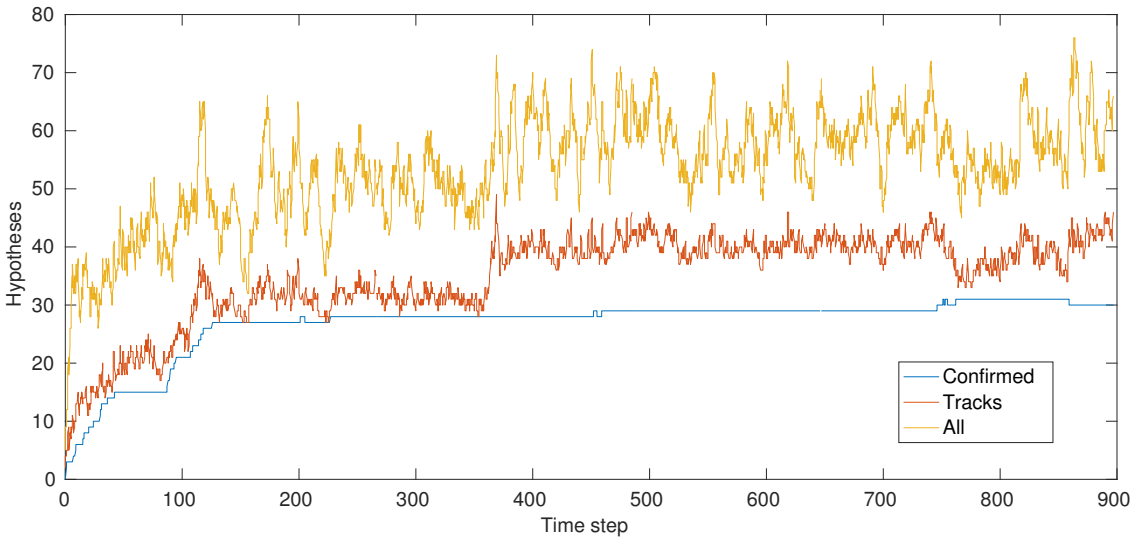


Figure 4.10: Number of maintained hypotheses by the HISP filter. The blue line corresponds to confirmed tracks, the red line marks the number of hypotheses that correspond to tracks, and the yellow one is the total number of hypotheses (including those used only for display).

Chapter 5

Conclusion

This thesis presents the concepts of recursive Bayesian state estimation, a powerful framework for propagating uncertainty and gaining information about processes measured with imperfect sensors, and proposes solutions to challenging problems using this framework. In doing so, novel estimation methods embedded in the Bayesian paradigm are introduced, and integrated with existing techniques to build solutions to different problems.

Chapter 2 discusses the background of recursive Bayesian state estimation. Bayes' filter, a conceptual method based on these ideas, was presented. This filter was shown to be intractable in the general case, although it sets the stage for many practical implementations, the most common of which were presented. Among these, the Kalman filter was shown to provide a statistically optimal solution when the propagated distributions are Gaussian, and linear measurement and dynamical models are used. Since it is also efficient computationally and easy to implement, many attempts have been made at obtaining an analogous filter in the non-linear non-Gaussian case. Two of these approaches were presented – the Extended Kalman filter, for instance, linearizes the models in order to approximate the propagated distributions with a Kalman update, while the Unscented Kalman filter approximates the distribution as a set of discrete samples, after which empirical statistics are obtained from the propagated distributions.

The sequential Monte Carlo family of filters was also shown, which does not make any assumptions on the form of the filtered distributions or of the dynamical or measurement models that are used. These are ideal for situations where making the assumption that a distribution is Gaussian or attempting to linearize the models incurs significant losses in estimation accuracy. The trade-off is that since this strategy is based on propagating samples of the distribution, it can become

computationally burdensome.

Based on the previously discussed concepts, a novel way to use the Kalman filter in non-linear problems was developed in this work. The core concept is that a Monte Carlo approach is used to map probability distributions into the extended sensor state space, where the measurement model becomes linear. Having done this, the distribution is approximated as a Gaussian in this new space, enabling the application of a linear Kalman update. After doing this, the same Monte Carlo approach is used to map the distribution back to the original space.

An application of the proposed single-object filter was to space situational awareness, shown in Chapter 3. In this domain, it is common for the focus to be on developing numerical propagation algorithms of high sophistication, but little attention is paid to the propagation of uncertainty. Here, the objects of interest are satellites or debris in Earth orbit which are observed with telescopes or radars, and the proxy space was a sensor-centered system of spherical co-ordinates. The filtering ideas that were applied in the previous chapter were applied to this domain, and it was shown that the filter performed well in propagating orbital uncertainty. An initial orbit determination method which generates a prior from a single measurement was shown, based on the energy constraints of orbiting objects, and the filter manages to propagate it in spite of its the non-Gaussian form. The prediction was done with a linear solution to the two-body problem, the Shepperd matrix. Filtering results were presented using realistic simulated data. Further work in this area can involve the inclusion of more sophisticated prediction models, exploring different ways to map distributions between spaces, and extending the algorithm to use measurements from a more diverse set of sensors.

Chapter 4 shows how a novel multi-object estimation framework can be used to produce an efficient multi-sensor multi-object estimation filter, using the single-object filter for space situational awareness that had been previously described. This is an ideal method for catalog maintenance, as it is normally of interest to keep track of objects in space, whether man-made or not, in order to safeguard space-based infrastructure. The HISP filter, a principled, efficient estimation algorithm which allows for a large degree of flexibility when modeling filtering problems, was used for this purpose. The resulting method was shown to perform well in tracking large numbers of orbiting targets while maintaining reasonable running times. As further work, it would be interesting to analyze how a parallel implementation of the HISP filter would perform in a scenario where thousands of targets are tracked, as this algorithm has a highly parallelizable structure and the problem of maintaining large catalogs and integrating data from sensors in different places along the surface

of the Earth is a real problem which this method could provide a solution for.

Appendix A

Testing for Multivariate Normality

MANY of the methods described in this thesis use the assumption that the underlying probability distributions that are being analyzed are Gaussian. This is due to several reasons: Gaussian distributions are simple but descriptive, making them ideal to model random processes; They have a compact representation, requiring only two parameters to be fully characterized; They have several properties which simplify computing the results of usual operations done on probability distributions; and natural processes seem to follow them often, which is not unreasonable considering that the central limit theorem dictates that a sum of random variables will converge to a Gaussian distribution as the number of summands grows to infinity.

In this appendix, a summary is given about methods to assess whether a set of samples plausibly originates from a Gaussian distribution, since some of the results presented on this thesis rely on approximating empirical distributions as continuous Gaussian ones. A departure from multivariate normality in this case will normally indicate that a measure of the information contained in the samples is being lost by the approximation, which would make the results of the methods that use it less reliable. In order to perform this assessment, two types of tools are usually available. So called graphical methods rely on visually inspecting quantile plots in order to assess how closely the empirical distribution resembles the theoretical Gaussian distribution, while hypothesis tests evaluate the likelihood of a hypothesis given the available data. Since graphical methods cannot be automated, the focus of this appendix will be on hypothesis testing.

A.1 Tests for univariate Gaussian distributions

A hypothesis test is a quantitative method which evaluates whether a given hypothesis is likely given the available data. A *null hypothesis* H_0 is tested against the *alternative hypothesis* H_a. The test then consists on evaluating a *test statistic*, which is a function of the available data, and evaluating its probability under the assumption that H_0 is true. The *p-value* is the probability of obtaining the test statistic, or a more extreme value, given that the null hypothesis is true. If this value is lower than a pre-specified confidence threshold α, the null hypothesis is *rejected* in favor of the alternative. If not, it is said that the available evidence is not enough to reject it. The performance of statistical tests is usually evaluated through their *power*. The power of a statistical test is defined as the probability of rejecting the null hypothesis given that the alternative is true. The higher the power of a test, the more sensitive it is to evidence in rejecting the null hypothesis [11].

In the univariate case (i.e., the dimension of the random variable is 1), commonly used tests to evaluate whether a sample comes from a Gaussian distribution are goodness of fit tests based on the theoretical cumulative distribution function, such as the Anderson-Darling test [3]; the Shapiro-Wilk test [84], which is generally regarded as having very high power to detect deviations from univariate normality [22]; and an examination of the third and fourth moments of the empirical distribution (skewness and kurtosis) [55].

A.2 Multivariate Gaussian tests

The tests outlined above cannot be in general extended to the multivariate case. For a random vector, having every one of its components be distributed Gaussian is a necessary, but not sufficient, condition for it to be distributed multivariate Gaussian [55]. Several statistical tests have designed to assess the validity of the assumption of multivariate normality, roughly following three different approaches: Those extending the Shapiro-Wilk test to the multivariate case, those assessing the multivariate skewness and kurtosis of the distribution, and those which are based on an analysis of the empirical characteristic function [61]. In the first category, tests such as the one proposed by Royston [78] apply a transformation based on the Shapiro-Wilk W test statistic to the random vectors which collapse them to a sample of univalued random variates, which can then be used to compute a W statistic. In the second category, the multivariate skewness and kurtosis are obtained for the

sample, for which the distribution is known in the case of a multivariate Gaussian generating distribution [59]. Tests exist for both skewness and kurtosis separately, but it has been shown that a test combining both has more power than testing individually [17]. Finally, the third category are based on a statistic which computes the distance between the empirical characteristic function of the obtained sample and the that of a Gaussian distribution. This class of tests extends the Epps-Pulley test for univariate normality [20], giving the name BHEP to this family of tests (Named after Baringhaus, Henze, Epps and Pulley) after its generalization to the multivariate case due to Baringhaus and Henze [6].

Monte Carlo studies have been used to evaluate the performance of these statistical tests, yielding a strong advantage for the BHEP family of tests [22, 61]. It also has the advantage of being invariant to linear transformations of the tested sample. For these reasons, it will be used in this work whenever it will be necessary to evaluate whether a random sample can reasonably be said to come from a Gaussian random variable or not. The Henze-Zirkler test [35] which generalizes the work by Baringhaus and Henze [6] can be seen below.

A.3 The Henze-Zirkler test

The characteristic function $\phi_{\boldsymbol{X}}(t)$ of a $d-$ dimensional random variable $\boldsymbol{X}$ is defined as

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{E}[e^{i\boldsymbol{t}'\boldsymbol{X}}] = \int e^{i\boldsymbol{t}'\boldsymbol{x}} d\boldsymbol{x}. \tag{A.1}$$

Distributions are uniquely characterized by their characteristic functions [11]. Based on this, the idea of the test is to compare the characteristic function of the empirical distribution of the samples $s = \{\boldsymbol{x}_i\}_{i=1}^n$

$$\psi_s(\boldsymbol{t}) = \frac{1}{N} \sum_{j=1}^{n} e^{i\boldsymbol{t}'\boldsymbol{x}_j} \tag{A.2}$$

with that of a Gaussian random variable,

$$e^{-\frac{1}{2}\|\boldsymbol{t}\|^2}. \tag{A.3}$$

To begin, the samples are normalized such that their mean is the zero vector and their covariance matrix is the identity matrix of appropriate dimension, obtaining a new set of samples $s'\,\{\boldsymbol{y}_i\}_{i=1}^n$. Then the test statistic is based on the weighted

difference between these two characteristic functions

$$D = \int \left| \psi_{s'}(\boldsymbol{t}) - e^{-\frac{1}{2}\|\boldsymbol{t}\|^2} \right|^2 \phi(\boldsymbol{t}) dt, \tag{A.4}$$

with some weighting function $\phi(t)$. The Henze-Zirkler test chooses this function to be a Gaussian kernel:

$$\phi(t) = (2\pi\beta^2)^{-d/2} e^{-\frac{\|\boldsymbol{t}\|^2}{2\beta^2}}, \tag{A.5}$$

with a given value for β which parametrizes the weighing function. This value can be chosen such that the mean square integration error is minimized, by making an analogy with kernel density estimation:

$$\beta_{\text{opt}} = \frac{1}{\sqrt{2}} \left[\frac{(2d+1)n}{4} \right]^{\frac{1}{d+4}}. \tag{A.6}$$

In [35] it is shown that (A.4) can be rewritten as

$$D = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-\frac{\beta^2}{2}\|\boldsymbol{y}_i - \boldsymbol{y}_j\|} - 2(1+\beta^2)^{-\frac{d}{2}} \frac{1}{n} \sum_{j=1}^{n} e^{-\frac{\beta^2}{2(1+\beta)^2}\|\boldsymbol{y}_j\|^2} + (1+2\beta^2)^{-\frac{d}{2}}. \tag{A.7}$$

The limiting distribution of the statistic $T = nD$ is not available in closed form, but it is approximated by Henze and Zirkler as a lognormal distribution $\ln\mathcal{N}(m, s^2)$ with mean

$$m = 1 - (1+2\beta^2)^{-\frac{d}{2}} \left[1 + \frac{d\beta^2}{1+2\beta^2} + \frac{d(d+2)\beta^4}{2(1+2\beta^2)^2} \right], \tag{A.8}$$

and variance

$$\begin{aligned} s^2 = 2(1+4\beta^2)^{-\frac{d}{2}} + 2(1+2\beta^2)^{-d} \left[1 + \frac{2d\beta^4}{(1+2\beta^2)^2} + \frac{3d(d+2)\beta^8}{4(1+2\beta^2)^4} \right] \\ - 4w(\beta)^{\frac{d}{2}} \left[1 + \frac{3d\beta^4}{2w(\beta)} + \frac{d(d+2)\beta^8}{2w(\beta)^2} \right], \end{aligned} \tag{A.9}$$

where $w(\beta) = (1+\beta^2)(1+3\beta^2)$. This way, the p-value of the test can be obtained as

$$p = 1 - \Phi^{-1}(nD; m, s^2), \tag{A.10}$$

with Φ^{-1} the inverse cumulative distribution function of a log-normal random variable. This p-value is then compared with a chosen confidence threshold in order to evaluate the likelihood of the samples coming from a multivariate Gaussian random variable.

Bibliography

[1] D. L. Alspach and H. W. Sorenson. Nonlinear bayesian estimation using gaussian sum approximations. *IEEE transactions on automatic control*, 17(4):439–448, 1972.

[2] Z. Altamimi, X. Collilieux, and L. Métivier. ITRF2008: an improved solution of the international terrestrial reference frame. *Journal of Geodesy*, 85(8):457–473, 2011.

[3] T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769, 1954.

[4] V. Bally, G. Pagès, and J. Printems. A stochastic quantization method for nonlinear problems. *Monte Carlo Methods and Applications*, 7:21–34, 2001.

[5] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press Professional, Inc., San Diego, CA, USA, 1987.

[6] L. Baringhaus and N. Henze. A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35:339–348, 1988.

[7] R. H. Battin. *An introduction to the mathematics and methods of astrodynamics*. AIAA, 1999.

[8] T. Byron, R. Schutz, and G. H. Born. *Statistical orbit determination*. Academic Press, 2004.

[9] G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[10] L. Combrinck. Satellite laser ranging. In *Sciences of Geodesy-I*, pages 301–338. Springer, 2010.

[11] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1962.

[12] P. Del Moral and J. Houssineau. Particle Association Measures and Multiple Target Tracking. In *Theoretical Aspects of Spatial-Temporal Modeling*, pages 1–30. Springer, 2015.

[13] E. D. Delande, C. Frueh, J. Franco, J. Houssineau, and D. E. Clark. A novel multi-object filtering approach for space situational awareness. *Journal of Guidance, Control, and Dynamics*, 2017. submitted.

[14] E. D. Delande, J. Houssineau, and D. E. Clark. Multi-object filtering with stochastic populations. *ArXiv e-prints*, January 2015.

[15] E. D. Delande, J. Houssineau, J. Franco, C. Frueh, and D. E. Clark. A new multi-target tracking algorithm for a large number of orbiting objects. In *Proceedings of the 27th AAS/AIAA Space Flight Mechanics Meeting*, San Antonio, TX, 2017.

[16] K. J. DeMars and M. K. Jah. Probabilistic initial orbit determination using Gaussian mixture models. *Journal of Guidance, Control, and Dynamics*, 36(5):1324–1335, 2013.

[17] J. A. Doornik and H. Hansen. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1):927–939, 2008.

[18] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*, volume 1. Springer-Verlag, 2001.

[19] Z. Duan, V. P. Jilkov, and X. Rong Li. State estimation with quantized measurements: Approximate MMSE approach. In *Information Fusion (FUSION), 2008 11th Conference on*, pages 1–6. IEEE, 2008.

[20] T. W. Epps and L. B. Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726, 1983.

[21] P. R. Escobal. Methods of orbit determination. *New York: Wiley, 1965*, 1965.

[22] P. J. Farrell, M. Salibian-Barrera, and K. Naczk. On tests for multivariate normality and associated simulation studies. *Journal of Statistical Computation and Simulation*, 77(12):1065–1080, 2007.

[23] J. Franco, E. D. Delande, C. Frueh, J. Houssineau, and D. E. Clark. A spherical co-ordinate space parameterisation for orbit estimation. In *Proceedings of the 2016 IEEE Aerospace Conference*, pages 1–12, 2016.

[24] J. Franco, E. D. Delande, C. Frueh, J. Joussineau, and D. Clark. Probabilistic orbit determination using a sensor co-ordinate parametrization. *Journal of Guidance, Control and Dynamics*, —(—), Under review.

[25] J. Franco, J. Houssineau, D. E. Clark, and C. Rickman. Simultaneous tracking of multiple particles and sensor position estimation in fluorescence microscopy images. In *Control, Automation and Information Sciences (ICCAIS), 2013 International Conference on*, pages 122–127. IEEE, 2013.

[26] C. Frueh, T. M. Kelecy, and M. K. Jah. Coupled orbit-attitude dynamics of high area-to-mass ratio (HAMR) objects: influence of solar radiation pressure, earth's shadow and the visibility in light curves. *Celestial Mechanics and Dynamical Astronomy*, 117(4):385–404, 2013.

[27] R. H. Gooding. A new procedure for the solution of the classical problem of minimal orbit determination from three lines of sight. *Celestial Mechanics and Dynamical Astronomy*, 66(4):387–423, 1996.

[28] W. H. Goodyear. Completely general closed-form solution for coordinates and partial derivative of the two-body problem. *Astronomical Journal*, 70(3):189 – 192, 1965.

[29] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

[30] M. S. Grewal and A. P. Andrews. Applications of Kalman filtering in aerospace, 1960 to the present, historical perspectives. *IEEE Control Systems*, 30(3):69–78, 2010.

[31] M. S. Grewal, L. R. Weill, and A. P. Andrews. *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, 2007.

[32] P. Gurfil and P. K. Seidelmann. *Celestial Mechanics and Astrodynamics: Theory and Practice*. Astrophysics and Space Science Library. Springer Berlin Heidelberg, 2016.

[33] O. Hagen, J. Houssineau, I. Schlangen, E. D. Delande, J. Franco, and D. E. Clark. Joint estimation of telescope drift and space object tracking. In *Aerospace Conference, 2016 IEEE*, pages 1–10. IEEE, 2016.

[34] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[35] N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617, 1990.

[36] Y. C. Ho and R. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9(4):333–339, 1964.

[37] J. Houssineau. *Representation and Estimation of Stochastic Populations*. PhD thesis, Heriot Watt University, 2015.

[38] J. Houssineau and D. E. Clark. Multi-target filtering with linearised complexity. *ArXiv e-prints*, 2016. arXiv:1404.7408v2.

[39] J. Houssineau, D. E. Clark, S. Ivekovic, C. S. Lee, and J. Franco. A unified approach for multi-object triangulation, tracking and camera calibration. *IEEE Transactions on Signal Processing*, 64(11):2934–2948, 2016.

[40] J. Houssineau and D. Laneuville. PHD filter with diffuse spatial prior on the birth process with applications to GM-PHD filter. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE, 2010.

[41] M. Hürzeler and H. R. Künsch. Monte carlo approximations for general state-space models. *Journal of Computational and graphical Statistics*, 7(2):175–193, 1998.

[42] Rob J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50, 05 1996.

[43] A. H. Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[44] B.A. Jones, S. Gehly, and P. Axelrad. Measurement-based birth model for a space object cardinalized probability hypothesis density filter. In *AIAA/AAS Astrodynamics Specialist Conference*, page 4311, 2014.

[45] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[46] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *American Control Conference, Proceedings of the 1995*, volume 3, pages 1628–1632 vol.3, Jun 1995.

[47] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[48] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[49] J. A. Kennewell and B. N. Vo. An overview of space situational awareness. In *Information Fusion (FUSION), 2013 16th International Conference on*, pages 1029–1036, July 2013.

[50] G. Kirchner, F. Koidl, F. Friederich, I. Buske, U. Völker, and W. Riede. Laser measurements to space debris from graz SLR station. *Advances in Space Research*, 51(1):21 – 24, 2013.

[51] U. Konda, P. Singla, T. Singh, and P. D. Scott. State uncertainty propagation in the presence of parametric uncertainty and additive white noise. *Journal of Dynamic Systems, Measurement, and Control*, 133(5):051009, 2011.

[52] C. S. Lee, D. E. Clark, and J. Salvi. SLAM with single cluster PHD filters. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2096–2101, May 2012.

[53] C. S. Lee, D. E. Clark, and J. Salvi. SLAM with dynamic targets via single-cluster PHD filtering. *Selected Topics in Signal Processing, IEEE Journal of*, 7(3):543–552, 2013.

[54] C. S. Lee, J. Franco, J. Houssineau, and D. E. Clark. Accelerating the single cluster PHD filter with a GPU implementation. In *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on*, pages 53–58. IEEE, Dec 2014.

[55] S. W. Looney. How to use tests for univariate normality to assess multivariate normality. *The American Statistician*, 49(1):64–70, 1995.

[56] R. P. S. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1152–1178, 2003.

[57] R. P. S. Mahler. PHD filters of higher order in target number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543, 2007.

[58] R. P. S. Mahler. *Statistical multisource-multitarget information fusion*, volume 685. Artech House, Boston, 2007.

[59] K. V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, pages 519–530, 1970.

[60] A. Mashiku, J. L. Garrison, and J. R. Carpenter. Statistical orbit determination using the particle filter for incorporating non-Gaussian uncertainties. In *AIAA/AAS Astrodynamics Specialist Conference*, pages 13–16, 2012.

[61] C. J. Mecklin and D. J. Mundfrom. A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, 75(2):93–107, 2005.

[62] R. M. et al Neal. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162, 2011.

[63] A. Pak, J. Correa, M. Adams, D. E. Clark, E. D. Delande, J. Houssineau, J. Franco, and C. Frueh. Joint target detection and tracking filter for Chilbolton advanced meteorological radar data processing. In *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2016.

[64] M. R. Pearlman, J. J. Degnan, and J. M. Bosworth. The international laser ranging service. *Advances in Space Research*, 30(2):135–143, 2002.

[65] M. Pereyra, P. Schniter, E. Chouzenoux, J. C. Pesquet, J. Y. Tourneret, A. O. Hero, and S. McLaughlin. A survey of stochastic simulation and optimization methods in signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):224–241, 2016.

[66] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

[67] J. E. Potter. Matrix quadratic solutions. *SIAM Journal on Applied Mathematics*, 14(3):496–501, 1966.

[68] C. Prada, J. Thomas, and M. Fink. The iterative time reversal process: Analysis of the convergence. *The Journal of the Acoustical Society of America*, 97(1):62–71, 1995.

[69] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. John Wiley and Sons, Inc., 2000.

[70] D. Raihan and S. Chakravorty. A UKF-PF based hybrid estimation scheme for space object tracking. *arXiv preprint arXiv:1409.7723*, 2014.

[71] D. B. Reid. An algorithm for tracking multiple targets. *Automatic Control, IEEE Transactions on*, 24(6):843–854, 1979.

[72] S. Reuter, B. Vo, B. Vo, and K. Dietmayer. The labeled multi-Bernoulli filter. *Signal Processing, IEEE Transactions on*, 62(12):3246–3260, 2014.

[73] B. Ristic, D. E. Clark, and N. Gordon. Calibration of multi-target tracking algorithms using non-cooperative targets. *Selected Topics in Signal Processing, IEEE Journal of*, 7(3):390–398, 2013.

[74] G. O. Roberts and R. L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[75] X. Rong Li and V. P. Jilkov. Survey of maneuvering target tracking: dynamic models. In *AeroSense 2000*, pages 212–235. International Society for Optics and Photonics, 2000.

[76] A. Rossi. The earth orbiting space debris. *Serbian Astronomical Journal*, 170:1–12, 2005.

[77] M. Roth, E. Özkan, and F. Gustafsson. A student's t filter for heavy tailed process and measurement noise. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5770–5774, May 2013.

[78] J. P. Royston. Some techniques for assessing multivarate normality based on the Shapiro-Wilk W. *Applied Statistics*, pages 121–133, 1983.

[79] F. Sanson and C. Frueh. Noise quantification in optical observations of resident space objects for probability of detection and likelihood. In *AAS/AIAA Astrodynamic Specialist Conference, Vail, CO*, pages 15–634, 2015.

[80] S. Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[81] I. Schlangen, E. D. Delande, J. Houssineau, and D. E. Clark. A PHD filter with negative binomial clutter. In *Information Fusion (FUSION), 2016 19th International Conference on*, pages 658–665. IEEE, 2016.

[82] I. Schlangen, J. Franco, J. Houssineau, W. T. E. Pitkeathly, D. E. Clark, I. Smal, and C. Rickman. Marker-less stage drift correction in super-resolution microscopy using the single-cluster PHD filter. *IEEE Journal of Selected Topics in Signal Processing*, 10(1):193–202, 2016.

[83] G. Seeber. *Satellite geodesy: foundations, methods, and applications*. Walter de Gruyter, 2003.

[84] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[85] S. W. Shepperd. Universal Keplerian state transition matrix. *Celestial Mechanics*, 35:129 – 144, 1985.

[86] C. Simpson, A. Hunter, S. Vorgul, E. D. Delande, J. Franco, and D. E. Clark. Likelihood modelling of the space geodesy facility laser ranging sensor for Bayesian filtering. In *Sensor Signal Processing for Defence (SSPD), 2016*, pages 1–5. IEEE, 2016.

[87] N. Singh, A. Poore, C. Sheaff, J. Aristoff, and M. K. Jah. Multiple Hypothesis Tracking (MHT) for Space Surveillance: Results and Simulation Studies. In *Advanced Maui Optical and Space Surveillance Technologies Conference*, page E16, September 2013.

[88] H. W. Sorenson and D. L. Alspach. Recursive bayesian estimation using gaussian sums. *Automatica*, 7(4):465–479, 1971.

[89] A. Swain and D. E. Clark. Extended object filtering using spatial independent cluster processes. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE, 2010.

[90] A. Swain and D. E. Clark. First-moment filters for spatial independent cluster processes. In *SPIE Defense, Security, and Sensing*, pages 76970I–76970I. International Society for Optics and Photonics, 2010.

[91] A. Swain and D. E. Clark. The single-group PHD filter: an analytic solution. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–8. IEEE, 2011.

[92] P. Tenenbaum, S. T. Bryson, H. Chandrasekaran, J. Li, E. Quintana, J. D. Twicken, and J. M. Jenkins. An algorithm for the fitting of planet models

to Kepler light curves. In *SPIE Astronomical Telescopes+ Instrumentation*, pages 77400J–77400J. International Society for Optics and Photonics, 2010.

[93] G. Terejanu, P. Singla, T. Singh, and P. D. Scott. Adaptive Gaussian sum filter for nonlinear bayesian estimation. *IEEE Transactions on Automatic Control*, 56(9):2151–2156, 2011.

[94] S. Thrun, W. Burgard, and D. Fox. *Probabilistic robotics*. MIT press, 2005.

[95] G. Tommei, A. Milani, and A. Rossi. Orbit determination of space debris: admissible regions. *Celestial Mechanics and Dynamical Astronomy*, 97(4):289–304, 2007.

[96] K. Vishwajeet, P. Singla, and M. K. Jah. Nonlinear uncertainty propagation for perturbed two-body orbits. *Journal of Guidance, Control, and Dynamics*, 37(5):1415–1425, 2014.

[97] B. Vo and W. Ma. The gaussian mixture probability hypothesis density filter. *Signal Processing, IEEE Transactions on*, 54(11):4091–4104, 2006.

[98] B. Vo, S. Singh, and A. Doucet. Sequential Monte Carlo methods for multi-target filtering with random finite sets. *IEEE Transactions on Aerospace and electronic systems*, 41(4):1224–1245, 2005.

[99] B. Vo and B. Vo. Labeled random finite sets and multi-object conjugate priors. *Signal Processing, IEEE Transactions on*, 61(13):3460–3475, 2013.

[100] B. Vo, B. Vo, and A. Cantoni. The cardinality balanced multi-target multi-Bernoulli filter and its implementations. *Signal Processing, IEEE Transactions on*, 57(2):409–423, 2009.

[101] W. L. Winston and J. B. Goldberg. *Operations research: applications and algorithms*. Duxbury press Boston, 2004.

[102] Z. Zhang, F. Yang, H. Zhang, Z. Wu, J. Chen, P. Li, and W. Meng. The use of laser ranging to measure space debris. *Research in Astronomy and Astrophysics*, 12(2):212, 2012.